

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
17 April 2003 (17.04.2003)

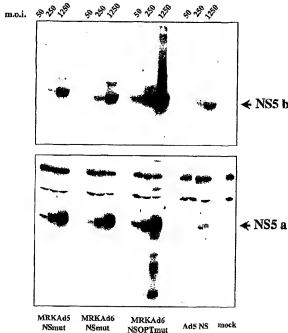
PCT

(10) International Publication Number
WO 03/031588 A2

- (51) International Patent Classification⁷: C12N
(21) International Application Number: PCT/US02/32512
(22) International Filing Date: 10 October 2002 (10.10.2002)
(25) Filing Language: English
(26) Publication Language: English
(30) Priority Data:
60/328,655 11 October 2001 (11.10.2001) US
60/363,774 13 March 2002 (13.03.2002) US
(71) Applicants (for all designated States except US): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI, S.P.A. [IT/IT]; Via Pontina KM. 30.600, 1-00040 POMEZIA (IT).
(72) Inventors; and
(75) Inventors/Applicants (for US only): EMINI, Emilio, A. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). KASLOW, David, C. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). BETT, Andrew, J. [CA/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). SHIVER, John, W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). NICOSIA, Alfredo [IT/IT]; Via Pontina KM. 30.600, 1-00040 Pomezia (IT). LAHM, Armin [DE/IT]; Via Pontina KM. 30.600, 1-00040 Pomezia (IT). LUZZAGO, Alessandra [IT/IT]; Via Pontina KM. 30.600, 1-00040 Pomezia (IT). CORTESE, Riccardo [IT/IT]; Via Pontina KM. 30.600, 1-00040 Pomezia (IT). COLLOCA, Stefano [IT/IT]; Via Pontina KM. 30.600, 1-00040 Pomezia (IT).
(74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

[Continued on next page]

(54) Title: HEPATITIS C VIRUS VACCINE



(57) Abstract: The present invention features Ad6 vectors and a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing an inactive NS5B RNA-dependent RNA polymerase region. The nucleic acid is particularly useful as a component of an adenovector or DNA plasmid vaccine providing a broad range of antigens for generating an HCV specific cell mediated immune (CMI) response against HCV.

Western blot on whole-cell extracts from HeLa cells infected at different multiplicity of infection (i.o.a.) indicated in the top with Adenovectors expressing the different HCV NS constructs. Mature NS5B and NS5A products were detected with specific antibodies.



(81) **Designated States (national):** AE, AG, AI, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TITLE OF THE INVENTION
HEPATITIS C VIRUS VACCINE

RELATED APPLICATIONS

- 5 The present application claims priority to provisional applications U.S. Serial No. 60/363,774, filed March 13, 2002, and U.S. Serial No. 60/328,655, filed October 11, 2001, each of which are hereby incorporated by reference herein.

BACKGROUND OF THE INVENTION

- 10 The references cited in the present application are not admitted to be prior art to the claimed invention.
- About 3% of the world's population are infected with the Hepatitis C virus (HCV). (Wasley *et al.*, *Semin. Liver Dis.* 20, 1-16, 2000.) Exposure to HCV results in an overt acute disease in a small percentage of cases, while in most
- 15 instances the virus establishes a chronic infection causing liver inflammation and slowly progresses into liver failure and cirrhosis. (Iwarson, *FEMS Microbiol. Rev.* 14, 201-204, 1994.) In addition, epidemiological surveys indicate an important role of HCV in the pathogenesis of hepatocellular carcinoma. (Kew, *FEMS Microbiol. Rev.* 14, 211-220, 1994, Alter, *Blood* 85, 1681-1695, 1995.)
- 20 Prior to the implementation of routine blood screening for HCV in 1992, most infections were contracted by inadvertent exposure to contaminated blood, blood products or transplanted organs. In those areas where blood screening of HCV is carried out, HCV is primarily contracted through direct percutaneous exposure to infected blood, *i.e.*, intravenous drug use. Less frequent methods of transmission
- 25 include perinatal exposure, hemodialysis, and sexual contact with an HCV infected person. (Alter *et al.*, *N. Engl. J. Med.* 341(8), 556-562, 1999, Alter, *J. Hepatol.* 31 Suppl. 88-91, 1999, *Semin. Liver Dis.* 201, 1-16, 2000.)
- The HCV genome consists of a single strand RNA about 9.5 kb encoding a precursor polyprotein of about 3000 amino acids. (Choo *et al.*, *Science* 244, 362-364, 1989, Choo *et al.*, *Science* 244, 359-362, 1989, Takamizawa *et al.*, *J. Virol.* 65, 1105-1113, 1991.) The HCV polyprotein contains the viral proteins in the order: C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B.
- 30

 Individual viral proteins are produced by proteolysis of the HCV polyprotein. Host cell proteases release the putative structural proteins C, E1, E2, and

p7, and create the N-terminus of NS2 at amino acid 810. (Mizushima *et al.*, *J. Virol.* 68, 2731-2734, 1994, Hijikata *et al.*, *P.N.A.S. USA* 90, 10773-10777, 1993.)

The non-structural proteins NS3, NS4A, NS4B, NS5A and NS5B presumably form the virus replication machinery and are released from the polypeptide. A zinc-dependent protease associated with NS2 and the N-terminus of NS3 is responsible for cleavage between NS2 and NS3. (Grakoui *et al.*, *J. Virol.* 67, 1385-1395, 1993, Hijikata *et al.*, *P.N.A.S. USA* 90, 10773-10777, 1993.) A distinct serine protease located in the N-terminal domain of NS3 is responsible for proteolytic cleavages at the NS3/NS4A, NS4A/NS4B, NS4B/NS5A and NS5A/NS5B junctions. (Bartenschlager *et al.*, *J. Virol.* 67, 3835-3844, 1993, Grakoui *et al.*, *Proc. Natl. Acad. Sci. USA* 90, 10583-10587, 1993, Tomei *et al.*, *J. Virol.* 67, 4017-4026, 1993.) NS4A provides a cofactor for NS3 activity. (Failla *et al.*, *J. Virol.* 68, 3753-3760, 1994, De Francesco *et al.*, U.S. Patent No. 5,739,002.)

NS5A is a highly phosphorylated protein conferring interferon resistance. (De Francesco *et al.*, *Semin. Liver Dis.*, 20(1), 69-83, 2000, Pawlitsky, *Viral Hepat. Suppl. 1*, 47-48, 1999.)

NS5B provides an RNA-dependent RNA polymerase. (De Francesco *et al.*, International Publication Number WO 96/37619, Behrens *et al.*, *EMBO* 15, 12-22, 1996, Lohmann *et al.*, *Virology* 249, 108-118, 1998.)

SUMMARY OF THE INVENTION

The present invention features Ad6 vectors and a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing an inactive NS5B RNA-dependent RNA polymerase region. The nucleic acid is particularly useful as a component of an adenovector or DNA plasmid vaccine providing a broad range of antigens for generating an HCV specific cell mediated immune (CMI) response against HCV.

A HCV specific CMI response refers to the production of cytotoxic T lymphocytes and T helper cells that recognize an HCV antigen. The CMI response may also include non-HCV specific immune effects.

Preferred nucleic acids encode a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide that is substantially similar to SEQ. ID. NO. 1 and has sufficient protease activity to process itself to produce at least a polypeptide substantially similar to the NS5B region present in SEQ. ID. NO. 1. The produced polypeptide corresponding to NS5B is enzymatically inactive. More preferably, the HCV polypeptide has sufficient

protease activity to produce polypeptides substantially similar to the NS3, NS4A, NS4B, NS5A, and NS5B regions present in SEQ. ID. NO. 1.

Reference to a "substantially similar sequence" indicates an identity of at least about 65% to a reference sequence. Thus, for example, polypeptides having an amino acid sequence substantially similar to SEQ. ID. NO. 1 have an overall amino acid identity of at least about 65% to SEQ. ID. NO. 1.

Polypeptides corresponding to NS3, NS4A, NS4B, NS5A, and NS5B have an amino acid sequence identity of at least about 65% to the corresponding region in SEQ. ID. NO. 1. Such corresponding polypeptides are also referred to herein as NS3, NS4A, NS4B, NS5A, and NS5B polypeptides.

Thus, a first aspect of the present invention describes a nucleic acid comprising a nucleotide sequence encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The encoded polypeptide has sufficient protease activity to process itself to produce an NS5B polypeptide that is enzymatically inactive.

In a preferred embodiment, the nucleic acid is an expression vector capable of expressing the Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide in a desired human cell. Expression inside a human cell has therapeutic applications for actively treating an HCV infection and for prophylactically treating against an HCV infection.

An expression vector contains a nucleotide sequence encoding a polypeptide along with regulatory elements for proper transcription and processing. The regulatory elements that may be present include those naturally associated with the nucleotide sequence encoding the polypeptide and exogenous regulatory elements not naturally associated with the nucleotide sequence. Exogenous regulatory elements such as an exogenous promoter can be useful for expression in a particular host, such as in a human cell. Examples of regulatory elements useful for functional expression include a promoter, a terminator, a ribosome binding site, and a polyadenylation signal.

Another aspect of the present invention describes a nucleic acid comprising a gene expression cassette able to express in a human cell a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The polypeptide can process itself to produce an enzymatically inactive NS5B protein. The gene expression cassette contains at least the following:

- a) a promoter transcriptionally coupled to a nucleotide sequence encoding a polypeptide;
- b) a 5' ribosome binding site functionally coupled to the nucleotide sequence,
- 5 c) a terminator joined to the 3' end of the nucleotide sequence, and
- d) a 3' polyadenylation signal functionally coupled to the nucleotide sequence.

Reference to "transcriptionally coupled" indicates that the promoter is positioned such that transcription of the nucleotide sequence can be brought about by RNA polymerase binding at the promoter. Transcriptionally coupled does not require that the sequence being transcribed is adjacent to the promoter.

10

Reference to "functionally coupled" indicates the ability to mediate an effect on the nucleotide sequence. Functionally coupled does not require that the coupled sequences be adjacent to each other. A 3' polyadenylation signal functionally coupled to the nucleotide sequence facilitates cleavage and polyadenylation of the transcribed RNA. A 5' ribosome binding site functionally coupled to the nucleotide sequence facilitates ribosome binding.

15

In preferred embodiments the nucleic acid is a DNA plasmid vector or an adenovector suitable for either therapeutic application in treating HCV or as an intermediate in the production of a therapeutic vector. Treating HCV includes actively treating an HCV infection and prophylactically treating against an HCV infection.

20

Another aspect of the present invention describes an adenovector comprising a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette able to express a polypeptide substantially similar to SEQ. ID. NO. 1 that is produced by a process involving (a) homologous recombination and (b) adenovector rescue. The homologous recombinant step produces an adenovirus genome plasmid. The adenovector rescue step produces the adenovector from the adenovirus genome plasmid.

25

Adenovirus genome plasmids described herein contain a recombinant adenovirus genome having a deletion in the E1 region and optionally in the E3 region and a gene expression cassette inserted into one of the deleted regions. The recombinant adenovirus genome is made of regions substantially similar to one or more adenovirus serotypes.

30

Another aspect of the present invention describes an adenovector consisting of the nucleic acid sequence of SEQ. ID. NO. 4 or a derivative thereof,

35

wherein said derivative thereof has the HCV polyprotein encoding sequence present in SEQ. ID. NO. 4 replaced with the HCV polyprotein encoding sequence of either SEQ. ID. NO. 3, SEQ. ID. NO. 10 or SEQ. ID. NO. 11.

Another aspect of the present invention describes a cultured

- 5 recombinant cell comprising a nucleic acid containing a sequence encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The recombinant cell has a variety of uses such as being used to replicate nucleic acid encoding the polypeptide in vector construction methods.

- Another aspect of the present invention describes a method of making
10 an adenovector comprising a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette able to express a polypeptide substantially similar to SEQ. ID. NO. 1. The method involves the steps of (a) producing an adenovirus genome plasmid containing a recombinant adenovirus genome with deletions in the E1 and E3 regions and a gene expression cassette inserted into one of the deleted regions and (b) rescuing the
15 adenovector from the adenovirus genome plasmid.

- Another aspect of the present invention describes a pharmaceutical composition comprising a vector for expressing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1 and a pharmaceutically acceptable carrier. The vector is suitable for administration and polypeptide
20 expression in a patient.

A "patient" refers to a mammal capable of being infected with HCV. A patient may or may not be infected with HCV. Examples of patients are humans and chimpanzees.

- Another aspect of the present invention describes a method of treating
25 a patient comprising the step of administering to the patient an effective amount of a vector expressing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ. ID. NO. 1. The vector is suitable for administration and polypeptide expression in the patient.

- The patient undergoing treatment may or may not be infected with
30 HCV. For a patient infected with HCV, an effective amount is sufficient to achieve one or more of the following effects: reduce the ability of HCV to replicate, reduce HCV load, increase viral clearance, and increase one or more HCV specific CMI responses. For a patient not infected with HCV, an effective amount is sufficient to achieve one or more of the following: an increased ability to produce one or more
35 components of a HCV specific CMI response to a HCV infection, a reduced

susceptibility to HCV infection, and a reduced ability of the infecting virus to establish persistent infection for chronic disease.

Another aspect of the present invention features a recombinant nucleic acid comprising an Ad6 region and a region not present in Ad6. Reference to
5 "recombinant" nucleic acid indicates the presence of two or more nucleic acid regions not naturally associated with each other. Preferably, the Ad6 recombinant nucleic acid contains Ad6 regions and a gene expression cassette coding for a polypeptide heterologous to Ad6.

Other features and advantages of the present invention are apparent
10 from the additional descriptions provided herein including the different examples. The provided examples illustrate different components and methodology useful in practicing the present invention. The examples do not limit the claimed invention. Based on the present disclosure the skilled artisan can identify and employ other components and methodology useful for practicing the present invention.

15

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B illustrate SEQ. ID. NO. 1.

Figures 2A, 2B, 2C, and 2D illustrate SEQ. ID. NO. 2. SEQ. ID. NO.
2 provides a nucleotide sequence coding for SEQ. ID. NO. 1 along with an optimized
20 internal ribosome entry site and TAAA termination. Nucleotides 1-6 provides an optimized internal ribosome entry site. Nucleotides 7-5961 code for a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide with nucleotides in positions 5137 to 5145 providing a AlaAlaGly sequence in amino acid positions 1711 to 1713 that renders NS5B inactive. Nucleotides 5962-5965 provide a TAAA termination.

Figures 3A, 3B, 3C, and 3D illustrate SEQ. ID. NO. 3. SEQ. ID. NO.
25 3 is a codon optimized version of SEQ. ID. NO. 2. Nucleotides 7-5961 encode a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide.

Figures 4A-4M illustrate MRKAd6-NSmut (SEQ. ID. NO. 4). SEQ.
ID. NO. 4 is an adenovector containing an expression cassette where the polypeptide
30 of SEQ. ID. NO. 1 is encoded by SEQ. ID. NO. 2. Base pairs 1-450 correspond to the Ad5 bp 1 to 450; base pairs 462 to 1252 correspond to the human CMV promoter; base pairs 1258 to 1267 correspond to the Kozak sequence; base pairs 1264 to 7222 correspond to the NS genes; base pairs 7231 to 7451 correspond to the BGH polyadenylation signal; base pairs 7469 to 9506 correspond to Ad5 base pairs 3511 to
35 5548; base pairs 9507 to 32121 correspond to Ad6 base pairs 5542 to 28156; base

pairs 32122 to 35117 correspond to Ad6 base pairs 30789 to 33784; and base pairs 35118 to 37089 correspond to Ad5 base pairs 33967 to 35935.

Figures 5A-5O illustrate SEQ. ID. NOs. 5 and 6. SEQ. ID. NO. 5 encodes a HCV Mct-NS3-NS4A-NS4B-NS5A-NS5B polypeptide with an active RNA dependent RNA polymerase. SEQ. ID. NO. 6 provides the amino acid sequence for the polypeptide.

Figures 6A-6C provide the nucleic acid sequence for pV1JnsA (SEQ. ID. NO. 7).

Figures 7A-7N provide the nucleic acid sequence for the Ad6 genome (SEQ. ID. NO. 8).

Figures 8A-8K provide the nucleic acid sequence for the Ad5 genome (SEQ. ID. NO. 9).

Figure 9 illustrates different regions of the Ad6 genome. The linear (37579 bp) ds DNA genome is indicated by two parallel lines and is divided into 100 map units. Transcription units are shown relative to their position and orientation in the genome. Early genes (E1A, E1B, E2A/B, E3 and E4 are indicated by gray arrows. Late genes (L1 to L5) , indicated by black arrows, are produced by alternative splicing of a transcript produced from the major late promoter (MLP) and all contain the tripartite leader (1, 2, 3) at their 5' ends. The E1 region is located from approximately 1.0 to 11.5 map units, the E2 region from 75.0 to 11.5 map units, E3 from 76.1 to 86.7 map units, and E4 from 99.5 to 91.2 map units. The major late transcription unit is located between 16.0 and 91.2 map units.

Figure 10 illustrates homologous recombination to recover pAdE1-E3+ containing Ad6 and Ad5 regions.

Figure 11 illustrates homologous recombination to recover a pAdE1-E3+ containing Ad6 regions.

Figure 12 illustrates a western blot on whole-cell extracts from 293 cells transfected with plasmid DNA expressing different HCV NS cassettes. Mature NS3 and NS5A products were detected with specific antibodies. "pV1Jns-NS" refers to a pV1JnsA plasmid where a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide is encoded by SEQ. ID. NO. 5, and SEQ. ID. NO. 5 is inserted between bases 1881 and 1912 of SEQ. ID. NO. 7. "pV1Jns-NSmut" refers to a pV1JnsA plasmid where SEQ. ID. NO. 2 is inserted between bases 1882 and 1925 of SEQ. ID. NO. 7. "pV1Jns-NSOPTmut" refers to a pV1JnsA plasmid where SEQ. ID. NO. 3 is inserted between bases 1881 and 1905 of SEQ. ID. NO. 7.

Figures 13A and 13B illustrate T cell responses by IFN γ ELISpot induced in C57black6 mice (A) and BalbC mice (B) by two injections of 25 μ g and 50 μ g, respectively, of plasmid DNA encoding the different HCV NS cassettes with Gene Electro-Transfer (GET).

5 Figure 14 illustrates protein expression from different adenovectors upon infection of HeLa cells. MRKAd5-NSmut is an adenovector based on an Ad5 sequence (SEQ. ID. NO. 9), where the Ad5 genome has an E1 deletion of base pairs 451 to 3510, an E3 deletion of base pairs 28134 to 30817, and has the NS3-NS4A-NS4B-NS5A-NS5B expression cassette as provided in base pairs 451 to 7468 of SEQ. ID. NO. 4 inserted between positions 450 and 3511. Ad5-NS is an adenovector based on an Ad5 backbone with an E1 deletion of base pairs 342 to 3523, and E3 deletion of base pairs 28134 to 30817 and containing an expression cassette encoding a NS3-NS4A-NS4B-NS5A-NS5B from SEQ. ID. NO. 5. "MRKAd6-NSOPTmut" refers to an adenovector having a modified SEQ. ID. NO. 4 sequence, wherein base pairs 1258
10 to 7222 of SEQ. ID. NO. 4 is replaced with SEQ. ID. NO. 3.

15 Figure 15 illustrates T cell responses by IFN γ ELISpot induced in C57black6 mice by two injections of 10^9 vp of adenovectors containing different HCV non-structural gene cassettes.

 Figures 16A-16D illustrate T cell responses by IFN γ ELISpot induced
20 in Rhesus monkeys by one or two injections of 10^{10} vp (A) or 10^{11} vp (B) of adenovectors containing different HCV non-structural gene cassettes.

 Figures 17A and 17B illustrates CD8+ T cell responses by IFN γ ICS induced in Rhesus monkeys by two injections of 10^{10} vp (A) or 10^{11} vp (B) of adenovectors encoding the different HCV non-structural gene cassettes.

25 Figures 18A-18F illustrate T cell responses by bulk CTL assay induced in Rhesus monkeys by two injections of 10^{11} vp of Ad5-NS (A), MRKAd5-NSmut (B), or MRKAd6-NSmut (C).

 Figure 19 illustrates the plasmid pE2.

 Figures 20A-D illustrates the partial codon optimized sequence
30 NSsuboptmut (SEQ. ID. NO. 10). Coding sequence for the Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide is from base 7 to 5961.

DETAILED DESCRIPTION OF THE INVENTION

The present invention features Ad6 vectors and nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide that contains an inactive NS5B region. Providing an inactive NS5B region supplies NS5B antigens while reducing the possibility of adverse side effects due to an active viral RNA polymerase. Uses of the featured nucleic acid include use as a vaccine component to introduce into a cell an HCV polypeptide that provides a broad range of antigens for generating a CMI response against HCV, and as an intermediate for producing such a vaccine component.

10 The adaptive cellular immune response can function to recognize viral antigens in HCV infected cells throughout the body due to the ubiquitous distribution of major histocompatibility complex (MHC) class I and II expression, to induce immunological memory, and to maintain immunological memory. These functions are attributed to antigen-specific CD4+ T helper (Th) and CD8+ cytotoxic T cells (CTL).

15 Upon activation via their specific T cell receptors, HCV specific Th cells fulfill a variety of immunoregulatory functions, most of them mediated by Th1 and Th2 cytokines. HCV specific Th cells assist in the activation and differentiation of B cells and induction and stimulation of virus-specific cytotoxic T cells. Together with CTL, Th cells may also secrete IFN- γ and TNF- α that inhibit replication and gene expression of several viruses. Additionally, Th cells and CTL, the main effector cells, can induce apoptosis and lysis of virus infected cells.

HCV specific CTL are generated from antigens processed by professional antigen presenting cells (pAPCs). Antigens can be either synthesized within or introduced into pAPCs. Antigen synthesis in a pAPC can be brought about by introducing into the cell an expression cassette encoding the antigen.

25 A preferred route of nucleic acid vaccine administration is an intramuscular route. Intramuscular administration appears to result in the introduction and expression of nucleic acid into somatic cells and pAPCs. HCV antigens produced in the somatic cells can be transferred to pAPCs for presentation in the context of MHC class I molecules. (Donnelly *et al.*, *Annu. Rev. Immunol.* 15:617-648, 1997.)

30 pAPCs process longer length antigens into smaller peptide antigens in the proteasome complex. The antigen is translocated into the endoplasmic reticulum/Golgi complex secretory pathway for association with MHC class I

proteins. CD8+ T lymphocytes recognize antigen associated with class I MHC via the T cell receptor (TCR) and the CD8 cell surface protein.

Using a nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide as a vaccine component allows for production of a broad range of
5 antigens capable of generating CMI responses from a single vector. The polypeptide should be able to process itself sufficiently to produce at least a region corresponding to NS5B. Preferred nucleic acids encode an amino acid sequence substantially similar to SEQ. ID. NO. 1 that has sufficient protease activity to process itself to produce
10 individual HCV polypeptides substantially similar to the NS3, NS4A, NS4B, NS5A, and NS5B regions present in SEQ. ID. NO. 1.

A polypeptide substantially similar to SEQ. ID. NO. 1 with sufficient protease activity to process itself in a cell provides the cell with T cell epitopes that are present in several different HCV strains. Protease activity is provided by NS3 and NS3/NS4A proteins digesting the Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide at
15 the appropriate cleavage sites to release polypeptides corresponding to NS3, NS4A, NS4B, NS5A, and NS5B. Self-processing of the Met-NS3-NS4A-NS4B-NS5A-NS5B generates polypeptides that approximate naturally occurring HCV polypeptides.

Based on the guidance provided herein a sufficiently strong immune response can be generated to achieve beneficial effects in a patient. The provided
20 guidance includes information concerning HCV sequence selection, vector selection, vector production, combination treatment, and administration.

I. HCV SEQUENCES

A variety of different nucleic acid sequences can be used as a vaccine
25 component to supply a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide to a cell or as an intermediate to produce vaccine components. The starting point for obtaining suitable nucleic acid sequences are preferably naturally occurring NS3-NS4A-NS4B-NS5A-NS5B polypeptide sequences modified to produce an inactive NS5B.

The use of a HCV nucleic acid sequence providing HCV non-structural antigens to generate a CMI response is mentioned by Cho *et al.*, *Vaccine* 17:1136-1144, 1999, Paliard *et al.*, International Publication Number WO 01/30812 (not admitted to be prior art to the claimed invention), and Coit *et al.*, International
35 Publication Number WO 01/38360 (not admitted to be prior art to the claimed invention). Such references fail to describe, for example, a polypeptide that processes

itself to produce an inactive NS5B, and the particular combinations of HCV sequences and delivery vehicles employed herein.

Modifications to a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide sequence can be produced by altering the encoding nucleic acid.

- 5 Alterations can be performed to create deletions, insertions and substitutions.

Small modifications can be made in NS5B to produce an inactive polymerase by targeting motifs essentially for replication. Examples of motifs critical for NS5B activity and modifications that can be made to produce an inactive NS5B are described by Lohmann *et al.*, *Journal of Virology* 71:8416-8426, 1997, and

- 10 Kolykhalov *et al.*, *Journal of Virology* 74:2046-2051, 2000.

Additional factors to take into account when producing modifications to a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide include maintaining the ability to self-process and maintaining T cell antigens. The ability of the HCV polypeptide to process itself is determined to a large extent by a functional NS3

15 protease. Modifications that maintain NS3 activity protease activity can be obtained by taking into account the NS3 protein, NS4A which serves as a cofactor for NS3, and NS3 protease recognition sites present within the NS3-NS4A-NS4B-NS5A-NS5B polypeptide.

- 20 Different modifications can be made to naturally occurring NS3-NS4A-NS4B-NS5A-NS5B polypeptide sequences to produce polypeptides able to elicit a broad range of T cell responses. Factors influencing the ability of a polypeptide to elicit a broad T cell response include the preservation or introduction of HCV specific T cell antigen regions and prevalence of different T cell antigen regions in different HCV isolates.

- 25 Numerous examples of naturally occurring HCV isolates are well known in the art. HCV isolates can be classified into the following six major genotypes comprising one or more subtypes: HCV-1/(1a,1b,1c), HCV-2/(2a,2b,2c), HCV-3/(3a,3b,10a), HCV-4/(4a), HCV-5/(5a) and HCV-6/(6a,6b,7b,8b,9a,11a). (Simmonds, *J. Gen. Virol.*, 693-712, 2001.) Examples of particular HCV sequences
- 30 such as HCV-BK, HCV-J, HCV-N, HCV-H, have been deposited in GenBank and described in various publications. (See, for example, Chamberlain *et al.*, *J. Gen. Virol.*, 1341-1347, 1997.)

- 35 HCV T cell antigens can be identified by, for example, empirical experimentation. One way of identifying T cell antigens involves generating a series of overlapping short peptides from a longer length polypeptide and then screening the

T-cell populations from infected patients for positive clones. Positive clones are activated/primed by a particular peptide. Techniques such as IFN γ -ELISPOT, IFN γ -Intracellular staining and bulk CTL assays can be used to measure peptide activity. Peptides thus identified can be considered to represent T-cell epitopes of the respective pathogen.

HCV T cell antigen regions from different HCV isolates can be introduced into a single sequence by, for example, producing a hybrid NS3-NS4A-NS4B-NS5A-NS5B polypeptide containing regions from two or more naturally occurring sequences. Such a hybrid can contain additional modifications, which preferably do not reduce the ability of the polypeptide to produce an HCV CMI response.

The ability of a modified Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide to process itself and produce a CMI response can be determined using techniques described herein or well known in the art. Such techniques include the use of IFN γ -ELISPOT, IFN γ -Intracellular staining and bulk CTL assays to measure a HCV specific CMI response.

A. Met-NS3-NS4A-NS4B-NS5A-NS5B Sequences

SEQ. ID. NO. 1 provides a preferred Met-NS3-NS4A-NS4B-NS5A-NS5B sequence. SEQ. ID. NO. 1 contains a large number of HCV specific T cell antigens that are present in several different HCV isolates. SEQ. ID. NO. 1 is similar to the NS3-NS4A-NS4B-NS5A-NS5B portion of the HCV BK strain nucleotide sequence (GenBank accession number M58335).

In SEQ. ID. NO. 1 anchor positions important for recognition by MHC class I molecules are conserved or represent conservative substitutions for 18 out of 20 known T-cell epitopes in the NS3-NS4A-NS4B-NS5A-NS5B portion of HCV polypeptides. With respect to the remaining two known T-cell epitopes, one has a non-conservative anchor substitution in SEQ. ID. NO. 1 that may still be recognized by a different HLA supertype and one epitope has one anchor residue not conserved. HCV T-cell epitopes are described in Chisari *et al.*, *Curr. Top. Microbiol. Immunol.*, 242:299-325, 2000, and Lechner *et al.* *J. Exp. Med.* 9:1499-1512, 2000.

Differences between the HCV-BK NS3-NS4A-NS4B-NS5A-NS5B nucleotide sequence and SEQ. ID. NO. 1 include the introduction of a methionine at the 5' end and the presence of modified NS5B active site residues in SEQ. ID. NO. 1.

The modification replaces GlyAspAsp with AlaAlaGly (residues 1711-1713) to inactivate NS5B.

The encoded HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide preferably has an amino acid sequence substantially similar to SEQ. ID. NO. 1. In different embodiments, the encoded HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide has an amino acid identity to SEQ. ID. NO. 1 of at least 65%, at least 75%, at least 85%, at least 95%, at least 99% or 100%; or differs from SEQ. ID. NO. 1 by 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, or 1-20 amino acids.

Amino acid differences between a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide and SEQ. ID. NO. 1 are calculated by determining the minimum number of amino acid modifications in which the two sequences differ. Amino acid modifications can be deletions, additions, substitutions or any combination thereof.

Amino acid sequence identity is determined by methods well known in the art that compare the amino acid sequence of one polypeptide to the amino acid sequence of a second polypeptide and generate a sequence alignment. Amino acid identity is calculated from the alignment by counting the number of aligned residue pairs that have identical amino acids.

Methods for determining sequence identity include those described by Schuler, G.D. in *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, Baxevanis, A.D. and Ouellette, B.F.F., eds., John Wiley & Sons, Inc., 2001; Yona, et al., in *Bioinformatics: Sequence, structure and databanks*, Higgins, D. and Taylor, W. eds, Oxford University Press, 2000; and *Bioinformatics: Sequence and Genome Analysis*, Mount, D.W., ed., Cold Spring Harbor Laboratory Press, 2001). Methods to determine amino acid sequence identity are codified in publicly available computer programs such as GAP (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.), BLAST (Altschul et al., *J. Mol. Biol.* 215(3):403-10, 1990), and FASTA (Pearson, *Methods in Enzymology* 183:63-98, 1990, R.F. Doolittle, ed.).

In an embodiment of the present invention sequence identity between two polypeptides is determined using the GAP program (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.). GAP uses the alignment method of Needleman and Wunsch. (Needleman, et al., *J. Mol. Biol.* 48:443-453, 1970.) GAP considers all possible alignments and gap positions between two sequences and creates a global alignment that maximizes the number of matched

residues and minimizes the number and size of gaps. A scoring matrix is used to assign values for symbol matches. In addition, a gap creation penalty and a gap extension penalty are required to limit the insertion of gaps into the alignment.

- 5 Default program parameters for polypeptide comparisons using GAP are the BLOSUM62 (Henikoff *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:10915-10919, 1992) amino acid scoring matrix (MATrix=blosum62.cmp), a gap creation parameter (GAPweight=8) and a gap extension parameter (LEnGhtweight=2).

- More preferred HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptides in addition to being substantially similar to SEQ. ID. NO. 1 across their entire length produce individual NS3, NS4A, NS4B, NS5A and NS5B regions that are substantially similar to the corresponding regions present in SEQ. ID. NO. 1. The corresponding regions in SEQ. ID. NO. 1 are provided as follows: Met-NS3 amino acids 1-632; NS4A amino acids 633-686; NS4B amino acids 687-947; NS5A amino acids 948-1394; and NS5B amino acids 1395-1985.

- 15 In different embodiments a NS3, NS4A, NS4B, NS5A and/or NS5B region has an amino acid identity to the corresponding region in SEQ. ID. NO. 1 of at least 65%, at least 75%, at least 85%, at least 95%, at least 99%, or 100%; or an amino acid difference of 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, or 1-20 amino acids.

- 20 Amino acid modifications to SEQ. ID. NO. 1 preferably maintain all or most of the T-cell antigen regions. Differences in naturally occurring amino acids are due to different amino acid side chains (R groups). An R group affects different properties of the amino acid such as physical size, charge, and hydrophobicity. Amino acids can be divided into different groups as follows: neutral and hydrophobic (alanine, valine, leucine, isoleucine, proline, tyryptophan, phenylalanine, and methionine); neutral and polar (glycine, serine, threonine, tryosine, cysteine, asparagine, and glutamine); basic (lysine, arginine, and histidine); and acidic (aspartic acid and glutamic acid).

- Generally, in substituting different amino acids it is preferable to exchange amino acids having similar properties. Substituting different amino acids within a particular group, such as substituting valine for leucine, arginine for lysine, and asparagine for glutamine are good candidates for not causing a change in polypeptide tertiary structure.

- Starting with a particular amino acid sequence and the known
35 degeneracy of the genetic code, a large number of different encoding nucleic acid

sequences can be obtained. The degeneracy of the genetic code arises because almost all amino acids are encoded by different combinations of nucleotide triplets or "codons". The translation of a particular codon into a particular amino acid is well known in the art (*see, e.g., Lewin GENES IV*, p. 119, Oxford University Press, 1990).

- 5 Amino acids are encoded by codons as follows:
A=Ala=Alanine: codons GCA, GCC, GCG, GCU
C=Cys=Cysteine: codons UGC, UGU
D=Asp=Aspartic acid: codons GAC, GAU
E=Glu=Glutamic acid: codons GAA, GAG
- 10 F=Phe=Phenylalanine: codons UUC, UUU
G=Gly=Glycine: codons GGA, GGC, GGG, GGU
H=His=Histidine: codons CAC, CAU
I=Ile=Isoleucine: codons AUA, AUC, AUU
K=Lys=Lysine: codons AAA, AAG
- 15 L=Leu=Leucine: codons UUA, UUG, CUA, CUC, CUG, CUU
M=Met=Methionine: codon AUG
N=Asn=Asparagine: codons AAC, AAU
P=Pro=Proline: codons CCA, CCC, CCG, CCU
Q=Gln=Glutamine: codons CAA, CAG
- 20 R=Arg=Arginine: codons AGA, AGG, CGA, CGC, CGG, CGU
S=Ser=Serine: codons AGC, AGU, UCA, UCC, UCG, UCU
T=Thr=Threonine: codons ACA, ACC, ACG, ACU
V=Val=Valine: codons GUA, GUC, GUG, GUU
W=Trp=Tryptophan: codon UGG
- 25 Y=Tyr=Tyrosine: codons UAC, UAU.

- Nucleic acid sequences can be optimized in an effort to enhance expression in a host. Factors to be considered include C:G content, preferred codons, and the avoidance of inhibitory secondary structure. These factors can be combined in different ways in an attempt to obtain nucleic acid sequences having enhanced
- 30 expression in a particular host. (See, for example, Donnelly *et al.*, International Publication Number WO 97/47358.)

- The ability of a particular sequence to have enhanced expression in a particular host involves some empirical experimentation. Such experimentation involves measuring expression of a prospective nucleic acid sequence and, if needed,
- 35 altering the sequence.

B. Encoding Nucleotide Sequences

SEQ. ID. NOs. 2 and 3 provide two examples of nucleotide sequences encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B sequence. The coding sequence of

- 5 SEQ. ID. NO. 2 is similar (99.4% nucleotide sequence identity) to the NS3-NS4A-NS4B-NS5A-NS5B region of the naturally occurring HCV-BK sequence (GenBank accession number M58335). SEQ. ID. NO. 3 is a codon-optimized version of SEQ. ID. NO. 2. SEQ. ID. NOs. 2 and 3 have a nucleotide sequence identity of 78.3%.

- 10 Differences between the HCV-BK NS3-NS4A-NS4B-NS5A-NS5B nucleotide (GenBank accession number M58335) and SEQ. ID. NO. 2, include SEQ. ID. NO. 2 having a ribosome binding site, an ATG methionine codon, a region coding for a modified NS5B catalytic domain, a TAAA stop signal and an additional 30 nucleotide differences. The modified catalytic domain codes for a AlaAlaGly (residues 1711-1713) instead of GlyAspAsp to inactivate NS5B.

- 15 A nucleotide sequence encoding a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide is preferably substantially similar to the SEQ. ID. NO. 2 coding region. In different embodiments, the nucleotide sequence encoding a HCV Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide has a nucleotide sequence identity to the SEQ. ID. NO. 2 coding region of at least 65%, at least 75%, at least 85%, at least 95%, at least 99%, or 100%; or differs from SEQ. ID. NO. 2 by 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, 1-20, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 nucleotides.

- 25 Nucleotide differences between a sequence coding Met-NS3-NS4A-NS4B-NS5A-NS5B and the SEQ. ID. NO. 2 coding region are calculated by determining the minimum number of nucleotide modifications in which the two sequences differ. Nucleotide modifications can be deletions, additions, substitutions or any combination thereof.

- 30 Nucleotide sequence identity is determined by methods well known in the art that compare the nucleotide sequence of one sequence to the nucleotide sequence of a second sequence and generate a sequence alignment. Sequence identity is determined from the alignment by counting the number of aligned positions having identical nucleotides.

- Methods for determining nucleotide sequence identity between two polynucleotides include those described by Schuler, in *Bioinformatics: A Practical*
35 *Guide to the Analysis of Genes and Proteins*, Baxeavanis, A.D. and Ouelette, B.F.F.,

eds., John Wiley & Sons, Inc, 2001; Yona *et al.*, in *Bioinformatics: Sequence, structure and databanks*, Higgins, D. and Taylor, W. eds, Oxford University Press, 2000; and *Bioinformatics: Sequence and Genome Analysis*, Mount, D.W., ed., Cold Spring Harbor Laboratory Press, 2001). Methods to determine nucleotide sequence identity are codified in publicly available computer programs such as GAP (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.), BLAST (Altschul *et al.*, *J. Mol. Biol.* 215(3):403-10, 1990), and FASTA (Pearson, W.R., *Methods in Enzymology* 183:63-98, 1990, R.F. Doolittle, ed.).

In an embodiment of the present invention, sequence identity between two polynucleotides is determined by application of GAP (Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, Wisc.). GAP uses the alignment method of Needleman and Wunsch. (Needleman *et al.*, *J. Mol. Biol.* 48:443-453, 1970.) GAP considers all possible alignments and gap positions between two sequences and creates a global alignment that maximizes the number of matched residues and minimizes the number and size of gaps. A scoring matrix is used to assign values for symbol matches. In addition, a gap creation penalty and a gap extension penalty are required to limit the insertion of gaps into the alignment. Default program parameters for polynucleotide comparisons using GAP are the nwsgapdna.cmp scoring matrix (MATrix=nwsgapdna.cmp), a gap creation parameter (GAPweight=50) and a gap extension parameter (LENGthweight=3).

More preferred HCV Met-NS3-NS4A-NS4B-NS5A-NS5B nucleotide sequences in addition to being substantially similar across its entire length, produce individual NS3, NS4A, NS4B, NS5A and NS5B regions that are substantially similar to the corresponding regions present in SEQ. ID. NO. 2. The corresponding coding regions in SEQ. ID. NO. 2 are provided as follows: Met-NS3, nucleotides 7-1902; NS4A nucleotides 1903-2064; NS4B nucleotides 2065-2847; NS5A nucleotides 2848-4188; NS5B nucleotides 4189-5661.

In different embodiments a NS3, NS4A, NS4B, NS5A and/or NS5B encoding region has a nucleotide sequence identity to the corresponding region in SEQ. ID. NO. 2 of at least 65%, at least 75%, at least 85%, at least 95%, at least 99% or 100%; or a nucleotide difference to SEQ. ID. NO. 2 of 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, 1-20, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 nucleotides.

C. Gene Expression Cassettes

A gene expression cassette contains elements needed for polypeptide expression. Reference to "polypeptide" does not provide a size limitation and includes protein. Regulatory elements present in a gene expression cassette generally include: (a) a promoter transcriptionally coupled to a nucleotide sequence encoding the polypeptide, (b) a 5' ribosome binding site functionally coupled to the nucleotide sequence, (c) a terminator joined to the 3' end of the nucleotide sequence, and (d) a 3' polyadenylation signal functionally coupled to the nucleotide sequence. Additional regulatory elements useful for enhancing or regulating gene expression or polypeptide processing may also be present.

Promoters are genetic elements that are recognized by an RNA polymerase and mediate transcription of downstream regions. Preferred promoters are strong promoters that provide for increased levels of transcription. Examples of strong promoters are the immediate early human cytomegalovirus promoter (CMV), and CMV with intron A. (Chapman *et al.*, *Nucl. Acids Res.* 19:3979-3986, 1991.) Additional examples of promoters include naturally occurring promoters such as the EF1 alpha promoter, the murine CMV promoter, Rous sarcoma virus promoter, and SV40 early/late promoters and the β -actin promoter; and artificial promoters such as a synthetic muscle specific promoter and a chimeric muscle-specific/CMV promoter (Li *et al.*, *Nat. Biotechnol.* 17:241-245, 1999, Hagstrom *et al.*, *Blood* 95:2536-2542, 2000).

The ribosome binding site is located at or near the initiation codon. Examples of preferred ribosome binding sites include CCACCAUGG, CCGCCAUGG, and ACCAUGG, where AUG is the initiation codon. (Kozak, *Cell* 44:283-292, 1986). Another example of a ribosome binding site is GCCACCAUGG (SEQ. ID. NO. 12).

The polyadenylation signal is responsible for cleaving the transcribed RNA and the addition of a poly (A) tail to the RNA. The polyadenylation signal in higher eukaryotes contains an AAUAAA sequence about 11-30 nucleotides from the polyadenylation addition site. The AAUAAA sequence is involved in signaling RNA cleavage. (Lewin, *Genes IV*, Oxford University Press, NY, 1990.) The poly (A) tail is important for the mRNA processing.

Polyadenylation signals that can be used as part of a gene expression cassette include the minimal rabbit β -globin polyadenylation signal and the bovine growth hormone polyadenylation (BGH). (Xu *et al.*, *Gene* 272:149-156, 2001, Post *et*

al., U.S. Patent U. S. 5,122,458.) Additional examples include the Synthetic Polyadenylation Signal (SPA) and SV40 polyadenylation signal. The SPA sequence is as follows: AAUAAAAGAUCUUUAUUUUAUUAAGAUCUGUGUGUUGGUUUUUUGUGUG (SEQ. ID. NO. 13).

- 5 Examples of additional regulatory elements useful for enhancing or regulating gene expression or polypeptide processing that may be present include an enhancer, a leader sequence and an operator. An enhancer region increases transcription. Examples of enhancer regions include the CMV enhancer and the SV40 enhancer. (Hitt *et al.*, *Methods in Molecular Genetics* 7:13-30, 1995, Xu, *et al.*,
10 *Gene* 272:149-156, 2001.) An enhancer region can be associated with a promoter.
- A leader sequence is an amino acid region on a polypeptide that directs the polypeptide into the proteasome. Nucleic acid encoding the leader sequence is 5' of a structural gene and is transcribed along the structural gene. An example of a leader sequences is tPA.
- 15 An operator sequence can be used to regulate gene expression. For example, the Tet operator sequence can be used to repress gene expression.

II. THERAPEUTIC VECTORS

- Nucleic acid encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B
20 polypeptide can be introduced into a patient using vectors suitable for therapeutic administration. Suitable vectors can deliver nucleic acid into a target cell without causing an unacceptable side effect.

- Cellular expression is achieved using a gene expression cassette encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide. The gene expression
25 cassette contains regulatory elements for producing and processing a sufficient amount of nucleic acid inside a target cell to achieve a beneficial effect.

- Examples of vectors that can be used for therapeutic applications include first and second generation adenovectors, helper dependent adenovectors, adeno-associated viral vectors, retroviral vectors, alpha virus vectors, Venezuelan
30 Equine Encephalitis virus vector, and plasmid vectors. (Hitt, *et al.*, *Advances in Pharmacology* 40:137-206, 1997, Johnston *et al.*, U.S. Patent No. 6,156,588, and Johnston *et al.*, International Publication Number WO 95/32733.) Preferred vectors for introducing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide into a subject are first generation adenoviral vectors and plasmid DNA vectors.

35

A. First Generation Adenovectors

First generation adenovector for expressing a gene expression cassette contain the expression cassette in an E1 and optionally E3 deleted recombinant adenovirus genome. The deletion in the E1 region is sufficiently large to remove elements needed for adenoviral replication.

First generation adenovectors for expressing a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide contain a E1 and E3 deleted recombinant adenovirus genome. The deletion in the E1 region is sufficiently large to remove elements needed for adenoviral replication. The combinations of deletions of the E1 and E3 regions are sufficiently large to accommodate a gene expression cassette encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide.

The adenovirus has a double-stranded linear genome with inverted terminal repeats at both ends. During viral replication, the genome is packaged inside a viral capsid to form a virion. The virus enters its target cell through viral attachment followed by internalization. (Hitt *et al.*, *Advances in Pharmacology* 40:137-206, 1997.)

Adenovectors can be based on different adenovirus serotypes such as those found in humans or animals. Examples of animal adenoviruses include bovine, porcine, chimp, murine, canine, and avian (CELO). Preferred adenovectors are based on human serotypes, more preferably Group B, C, or D serotypes. Examples of human adenovirus Group B, C, D, or E serotypes include types 2 ("Ad2"), 4 ("Ad4"), 5 ("Ad5"), 6 ("Ad6"), 24 ("Ad24"), 26 ("Ad26"), 34 ("Ad34") and 35 ("Ad35"). Adenovectors can contain regions from a single adenovirus or from two or more adenovirus.

In different embodiments adenovectors are based on Ad5, Ad6, or a combination thereof. Ad5 is described by Chroboczek, *et al.*, *J. Virology* 186:280-285, 1992. Ad6 is described in Figures 7A-7N. An Ad6 based vector containing Ad5 regions is described in the Example section provided below.

Adenovectors do not need to have their E1 and E3 regions completely removed. Rather, a sufficient amount the E1 region is removed to render the vector replication incompetent in the absence of the E1 proteins being supplied in *trans*; and the E1 deletion or the combination of the E1 and E3 deletions are sufficiently large enough to accommodate a gene expression cassette.

E1 deletions can be obtained starting at about base pair 342 going up to about base pair 3523 of Ad5, or a corresponding region from other adenoviruses.

Preferably, the deleted region involves removing a region from about base pair 450 to about base pair 3511 of Ad5, or a corresponding region from other adenoviruses. Larger E1 region deletions starting at about base pair 341 removes elements that facilitate virus packaging.

- 5 E3 deletions can be obtained starting at about base pair 27865 to about base pair 30995 of Ad5, or the corresponding region of other adenovectors. Preferably the deletion region involves removing a region from about base pair 28134 up to about base pair 30817 of Ad5, or the corresponding region of other adenovectors.

- 10 The combination of deletions to the E1 region and optionally the E3 region should be sufficiently large so that the overall size of the recombinant genome containing the gene expression cassette does not exceed about 105% of the wild type adenovirus genome. For example, as recombinant adenovirus Ad5 genomes increase size above about 105% the genome becomes unstable. (Bett *et al.*, *Journal of*
15 *Virology* 67:5911-5921, 1993.)

- Preferably, the size of the recombinant adenovirus genome containing the gene expression cassette is about 85% to about 105% the size of the wild type adenovirus genome. In different embodiments, the size of the recombinant adenovirus genome containing the expression cassette is about 100% to about
20 105.2%, or about 100%, the size of the wild type genome.

Approximately 7,500 kb can be inserted into an adenovirus genome with a E1 and E3 deletion. Without any deletion, the Ad5 genome is 35,935 base pairs and the Ad6 genome is 35,759 base pairs.

- Replication of first generation adenovectors can be performed by
25 supplying the E1 gene products *in trans*. The E1 gene product can be supplied *in trans*, for example, by using cell lines that have been transformed with the adenovirus E1 region. Examples of cells and cell lines transformed with the adenovirus E1 region are HEK 293 cells, 911 cells, PERC.6™ cells, and transfected primary human aminocytes cells. (Graham *et al.*, *Journal of Virology* 36:59-72, 1977, Schiedner *et al.*, *Human Gene Therapy* 11:2105-2116, 2000, Fallaux *et al.*, *Human Gene Therapy*
30 9:1909-1917, 1998, Bout *et al.*, U.S. Patent No. 6,033,908.)

- A Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette should be inserted into a recombinant adenovirus genome in the region corresponding to the deleted E1 region or the deleted E3 region. The expression cassette can have a
35 parallel or anti-parallel orientation. In a parallel orientation the transcription direction

of the inserted gene is the same direction as the deleted E1 or E3 gene. In an anti-parallel orientation transcription the opposite strand serves as a template and the transcription direction is in the opposite direction.

- In an embodiment of the present invention the adenovector has a gene expression cassette inserted in the E1 deleted region. The vector contains:
- 5 a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
 - b) a gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to the first region;
 - 10 c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
 - d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
 - 15 e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the third region; and
 - f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6 joined to the fourth region.
 - 20

In another embodiment of the present invention the adenovector has an expression cassette inserted in the E3 deleted region. The vector contains:

- 25 a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the first region;
- c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
- 30 d) a gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to the third region;

- c) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region.

In preferred different embodiments concerning adenovirus regions that are present: (1) the first, second, third, fourth, and fifth region corresponds to Ad5; (2) the first, second, third, fourth, and fifth region corresponds to Ad6; and (3) the first region corresponds to Ad5, the second region corresponds to Ad5, the third region corresponds to Ad6, the fourth region corresponds to Ad6, and the fifth region corresponds to Ad5.

B. DNA Plasmid Vectors

DNA vaccine plasmid vectors contain a gene expression cassette along with elements facilitating replication and preferably vector selection. Preferred elements provide for replication in non-mammalian cells and a selectable marker. The vectors should not contain elements providing for replication in human cells or for integration into human nucleic acid.

The selectable marker facilitates selection of nucleic acids containing the marker. Preferred selectable markers are those that confer antibiotic resistance. Examples of antibiotic selection genes include nucleic acid encoding resistance to ampicillin, neomycin, and kanamycin.

Suitable DNA vaccine vectors can be produced starting with a plasmid containing a bacterial origin of replication and a selectable marker. Examples of bacterial origins of replication providing for higher yields include the ColE1 plasmid-derived bacterial origin of replication. (Donnelly *et al.*, *Annu. Rev. Immunol.* 15:617-648, 1997.)

The presence of the bacterial origin of replication and selectable marker allows for the production of the DNA vector in a bacterial strain such as *E. coli*. The selectable marker is used to eliminate bacteria not containing the DNA vector.

III. AD6 RECOMBINANT NUCLEIC ACID

Ad6 recombinant nucleic acid comprises an Ad6 region substantially similar to an Ad6 region found in SEQ. ID. NO. 8, and a region not present in Ad6 nucleic acid. Recombinant nucleic acid comprising Ad6 regions have different uses such as in producing different Ad6 regions, as intermediates in the production of Ad6 based vectors, and as a vector for delivering a recombinant gene.

As depicted in Figure 9, the genomic organization of Ad6 is very similar to the genomic organization of Ad5. The homology between Ad5 and Ad6 is approximately 98%.

In different embodiments, the Ad6 recombinant nucleic acid comprises a nucleotide region substantially similar to E1A, E1B, E2B, E2A, E3, E4, L1, L2, L3, or L4, or any combination thereof. A substantially similar nucleic acid region to an Ad6 region has a nucleotide sequence identity of at least 65%, at least 75%, at least 85%, at least 95%, at least 99% or 100%; or a nucleotide difference of 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 1-17, 1-18, 1-19, 1-20, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 nucleotides. Techniques and embodiments for determining substantially similar nucleic acid sequences are described in Section I.B. *supra*.

Preferably, the recombinant Ad6 nucleic acid contains an expression cassette coding for a polypeptide not found in Ad6. Examples of expression cassettes include those coding for HCV regions and those coding for other types of polypeptides.

Different types of adenoviral vectors can be produced incorporating different amounts of Ad6, such as first and second generation adenovectors. As noted in Section II.A. *supra*, first generation adenovectors are defective in E1 and can replicate when E1 is supplied *in trans*.

Second generation adenovectors contain less adenoviral genome than first generation vectors and can be used in conjugation with complementing cell lines and/or helper vectors supplying adenoviral proteins. Second generation adenovectors are described in different references such as Russell, *Journal of General Virology* 81:2573-2604, 2000; Hitt *et al.*, 1997, Human Ad vectors for Gene Transfer, Advances in Pharmacology, Vol 40 Academic Press.

In an embodiment of the present invention, the Ad6 recombinant nucleic acid is an adenovirus vector defective in E1 that is able to replicate when E1 is

supplied *in trans*. Expression cassettes can be inserted into a deleted E1 region and/or a deleted E3 region.

An example of an Ad6 based adenoviral vector with an expression cassette provided in a deleted E1 region comprises or consists of:

- 5 a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
 - b) a gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to the first region;
 - c) a second adenovirus region from about base pair 3511 to about
10 base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
 - d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
 - 15 e) an optionally present fourth region from about base pair 28134 to about base pair 30817 corresponding to Ad5, or from about base pair 28157 to about base pair 30788 corresponding to Ad6, joined to the third region;
 - f) a fifth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base
20 pair 33784 corresponding to Ad6, wherein the fifth region is joined to the fourth region if the fourth region is present, or the fifth is joined to the third region if the fourth region is not present; and
 - g) a sixth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base
25 pair 35759 corresponding to Ad6, joined to the fifth region;
- wherein at least one Ad6 region is present.

In different embodiments of the invention, all of the regions are from Ad6; all of the regions except for the first and second are from Ad6; and 1, 2, 3, or 4 regions selected from the second, third, fourth, and fifth regions are from Ad6.

- 30 An example of an Ad6 based adenoviral vector with an expression cassette provided in a deleted E3 region comprises or consists of:

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the first region;
- c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
- d) a gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to the third region;
- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region;
- wherein at least one Ad6 region is present.

In different embodiment of the invention, all of the regions are from Ad6; all of the regions except for the first and second are from Ad6; and 1, 2, 3, or 4 regions selected from the second, third, fourth and fifth regions are from Ad6.

20

IV. VECTOR PRODUCTION

Vectors can be produced using recombinant nucleic acid techniques such as those involving the use of restriction enzymes, nucleic acid ligation, and homologous recombination. Recombinant nucleic acid techniques are well known in the art. (Ausubel, *Current Protocols in Molecular Biology*, John Wiley, 1987-1998, and Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press, 1989.)

Intermediate vectors are used to derive a therapeutic vector or to transfer an expression cassette or portion thereof from one vector to another vector. Examples of intermediate vectors include adenovirus genome plasmids and shuttle vectors.

30

Useful elements in an intermediate vector include an origin of replication, a selectable marker, homologous recombination regions, and convenient restriction sites. Convenient restriction sites can be used to facilitate cloning or release of a nucleic acid sequence.

Homologous recombination regions provide nucleic acid sequence regions that are homologous to a target region in another nucleic acid molecule. The homologous regions flank the nucleic acid sequence that is being inserted into the target region. In different embodiments homologous regions are preferably about 150 to 600 nucleotides in length, or about 100 to 500 nucleotides in length.

An embodiment of the present invention describes a shuttle vector containing a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette, a selectable marker, a bacterial origin of replication, a first adenovirus homology region and a second adenovirus homologous region that target the expression cassette to insert in or replace an E1 region. The first and second homology regions flank the expression cassette. The first homology region contains at least about 100 base pairs substantially homologous to at least the right end (3' end) of a wild-type adenovirus region from about base pairs 4-450. The second homology contains at least about 100 base pairs substantially homologous to at least the left end (5' end) of Ad5 from about base pairs 3511-5792, or the corresponding region from another adenovirus.

Reference to "substantially homologous" indicates a sufficient degree of homology to specifically recombine with a target region. In different embodiments substantially homologous refers to at least 85%, at least 95%, or 100% sequence identity. Sequence identity can be calculated as described in Section I.B. *supra*.

One method of producing adenovectors is through the creation of an adenovirus genome plasmid containing an expression cassette. The pre-Adenovirus plasmid contains all the adenovirus sequences needed for replication in the desired complementing cell line. The pre-Adenovirus plasmid is then digested with a restriction enzyme to release the viral ITR's and transfected into the complementing cell line for virus rescue. The ITR's must be released from plasmid sequences to allow replication to occur. Adenovector rescue results in the production of an adenovector containing the expression cassette.

A. Adenovirus Genome Plasmids

Adenovirus genome plasmids contain an adenovector sequence inside a longer-length plasmid (which may be a cosmid). The longer-length plasmid may contain additional elements such as those facilitating growth and selection in eukaryotic or bacterial cells depending upon the procedures employed to produce and maintain the plasmid. Techniques for producing adenovirus genome plasmids include those involving the use of shuttle vectors and homologous recombination, and those

involving the insertion of a gene expression cassette into an adenovirus cosmid. (Hitt *et al.*, *Methods in Molecular Genetics* 7:13-30, 1995, Danthinne *et al.*, *Gene Therapy* 7:1707-1714, 2000.)

- Adenovirus genome plasmids preferably have a gene expression cassette inserted into a E1 or E3 deleted region. In an embodiment of the present invention, the adenovirus genome plasmid contains a gene expression cassette inserted in the E1 deleted region, an origin of replication, a selectable marker, and the recombinant adenovirus region is made up of:
- a) a first adenovirus region from about base pair 1 to about base 450 corresponding to either Ad5 or Ad6;
 - b) a gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to the first region;
 - c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
 - d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
 - e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the third region;
 - f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region, and
 - g) an optionally present E3 region corresponding to all or part of the E3 region present in Ad5 or Ad6, which may be present for smaller inserts taking into account the overall size of the desired adenovector.

In another embodiment of the present invention the recombinant adenovirus genome plasmid has the gene expression cassette inserted in the E3 deleted region. The vector contains an origin of replication, a selectable marker, and the following:

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the expression cassette;
- c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to the second region;
- d) the gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to the third region;
- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region.

- 15 In different embodiments concerning adenovirus regions that are present: (1) the first, second, third, fourth, and fifth region corresponds to Ad5; (2) the first, second, third, fourth, and fifth region corresponds to Ad6; and (3) the first region corresponds to Ad5, the second region corresponds to Ad5, the third region corresponds to Ad6, the fourth region corresponds to Ad6, and the fifth region
- 20 corresponds to Ad5.

- An embodiment of the present invention describes a method of making an adenovector involving a homologous recombination step to produce a adenovirus genome plasmid and an adenovirus rescue step. The homologous recombination step involves the use of a shuttle vector containing a Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette flanked by adenovirus homology regions. The adenovirus homology regions target the expression cassette into either the E1 or E3 deleted region.
- 25

- In an embodiment of the present invention concerning the production of an adenovirus genome plasmid, the gene expression cassette is inserted into a vector comprising: a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6; a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to the second region; a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6,
- 30
- 35

joined to the second region; a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to the third region; and a fifth adenovirus region from about 33967 to about 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region. The adenovirus genome plasmid should contain an origin of replication and a selectable marker, and may contain all or part of the Ad5 or Ad6 E3 region.

In different embodiments concerning adenovirus regions that are present: (1) the first, second, third, fourth, and fifth region corresponds to Ad5; (2) the first, second, third, fourth, and fifth region corresponds to Ad6; and (3) the first region corresponds to Ad5, the second region corresponds to Ad5, the third region corresponds to Ad6, the fourth region corresponds to Ad6, and the fifth region corresponds to Ad5.

15 B. Adenovector Rescue

An adenovector can be rescued from a recombinant adenovirus genome plasmid using techniques known in the art or described herein. Examples of techniques for adenovirus rescue well known in the art are provided by Hitt *et al.*, *Methods in Molecular Genetics* 7:13-30, 1995, and Danthinne *et al.*, *Gene Therapy* 7:1707-1714, 2000.

A preferred method of rescuing an adenovector described herein involves boosting adenoviral replication. Boosting adenoviral replication can be performed, for example, by supplying adenoviral functions such as E2 proteins (polymerase, pre-terminal protein and DNA binding protein) as well as E4 orf6 on a separate plasmid. Example 10 *infra* illustrates the boosting of adenoviral replication to rescue an adenovector containing a codon optimized Met-NS3-NS4A-NS4B-NS5A-NS5B expression cassette.

V. PARTIAL-OPTIMIZED HCV ENCODING SEQUENCES

Partial optimization of HCV polyprotein encoding nucleic acid provides for a lesser amount of codons optimized for expression in a human than complete optimization. The overall objective is to provide the benefits of increased expression due to codon optimization, while facilitating the production of an adenovector containing HCV polyprotein encoding nucleic acid having optimized codons.

Complete optimization of an HCV polyprotein encoding sequence provides the most frequently observed human codon for each amino acid. Complete optimization can be performed using codon frequency tables well known in the art and using programs such as the BACKTRANSLATE program (Wisconsin Package version 10, Genetics Computer Group, GCG, Madison, Wisc.).

Partial optimization can be performed on an entire HCV polyprotein encoding sequence that is present (*e.g.*, NS3-NS5B), or one or more local regions that are present. In different embodiments the GC content for the entire HCV encoded polyprotein that is present is no greater than at least about 65%; and the GC content for one or more local regions is no greater than about 70%.

Local regions are regions present in HCV encoding nucleic acid, and can vary in size. For example, local regions can be about 60, about 70, about 80, about 90 or about 100 nucleotides in length.

Partial optimization can be achieved by initially constructing an HCV encoding polyprotein sequence to be partially optimized based on a naturally occurring sequence. Alternatively, an optimized HCV encoding sequence can be used as basis of comparison to produce a partial optimized sequence.

VI. HCV COMBINATION TREATMENT

The HCV Met-NS3-NS4A-NS4B-NS5A-NS5B vaccine can be used by itself to treat a patient, can be used in conjunction with other HCV therapeutics, and can be used with agents targeting other types of diseases. Additional therapeutics include additional therapeutic agents to treat HCV and diseases having a high prevalence in HCV infected persons. Agents targeting other types of disease include vaccines directed against HIV and HBV.

Additional therapeutics for treating HCV include vaccines and non-vaccine agents. (Zein, *Expert Opin. Investig. Drugs* 10:1457-1469, 2001.) Examples of additional HCV vaccines include vaccines designed to elicit an immune response against an HCV core antigen and the HCV E1, E2 or p7 region. Vaccine components can be naturally occurring HCV polypeptides, HCV mimotope polypeptides or nucleic acid encoding such polypeptides.

HCV mimotope polypeptides contain HCV epitopes, but have a different sequence than a naturally occurring HCV antigen. A HCV mimotope can be fused to a naturally occurring HCV antigen. References describing techniques for producing mimotopes in general and describing different HCV mimotopes are

provided in Felici *et al.* U.S. Patent No. 5,994,083 and Nicosia *et al.*, International Application Number WO 99/60132.

VII. PHARMACEUTICAL ADMINISTRATION

5 HCV vaccines can be formulated and administered to a patient using the guidance provided herein along with techniques well known in the art. Guidelines for pharmaceutical administration in general are provided in, for example, *Modern Vaccinology*, Ed. Kurstak, Plenum Med. Co. 1994; *Remington's Pharmaceutical Sciences 18th Edition*, Ed. Gennaro, Mack Publishing, 1990; and *Modern*
10 *Pharmaceutics 2nd Edition*, Eds. Banker and Rhodes, Marcel Dekker, Inc., 1990, each of which are hereby incorporated by reference herein.

HCV vaccines can be administered by different routes such intravenous, intraperitoneal, subcutaneous, intramuscular, intradermal, impression through the skin, or nasal. A preferred route is intramuscular.

15 Intramuscular administration can be preformed using different techniques such as by injection with or without one or more electric pulses. Electric mediated transfer can assist genetic immunization by stimulating both humoral and cellular immune responses.

Vaccine injection can be performed using different techniques, such as
20 by employing a needle or a needleless injection system. An example of a needleless injection system is a jet injection device. (Donnelly *et al.*, International Publication Number WO 99/52463.)

A. Electrically Mediated Transfer

25 Electrically mediated transfer or Gene Electro-Transfer (GET) can be performed by delivering suitable electric pulses after nucleic acid injection. (See Mathiesen, International Publication Number WO 98/43702). Plasmid injection and electroporation can be performed using stainless needles. Needles can be used in couples, triplets or more complex patterns. In one configuration the needles are
30 soldered on a printed circuit board that is a mechanical support and connects the needles to the electrical field generator by means of suitable cables.

The electrical stimulus is given in the form of electrical pulses. Pulses can be of different forms (square, sinusoidal, triangular, exponential decay) and different polarity (monopolar of positive or negative polarity, bipolar). Pulses can be
35 delivered either at constant voltage or constant current modality.

Different patterns of electric treatment can be used to introduce nucleic acid vaccines including HCV and other nucleic acid vaccines into a patient. Possible patterns of electric treatment include the following:

- Treatment 1: 10 trains of 1000 square bipolar pulses delivered every other second, pulse length 0.2 msec/phase, frequency 1000 Hz, constant voltage mode, 45 Volts/phase, floating current.
- Treatment 2: 2 trains of 100 square bipolar pulses delivered every other second, pulse length 2 msec/phase, frequency 100 Hz, constant current mode, 100 mA/phase, floating voltage.
- Treatment 3: 2 trains of bipolar pulses at a pulse length of about 2 msec/phase, for a total length of about 3 seconds, where the actual current going through the tissue is fixed at about 50 mA.

- Electric pulses are delivered through an electric field generator. A suitable generator can be composed of three independent hardware elements assembled in a common chassis and driven by a portable PC which runs the driving program. The software manages both basic and accessory functions. The elements of the device are: (1) signal generator driven by a microprocessor, (2) power amplifier and (3) digital oscilloscope.

- The signal generator delivers signals having arbitrary frequency and shape in a given range under software control. The same software has an interactive editor for the waveform to be delivered. The generator features a digitally controlled current limiting device (a safety feature to control the maximal current output). The power amplifier can amplify the signal generated up to +/- 150 V. The oscilloscope is digital and is able to sample both the voltage and the current being delivered by the amplifier.

B. Pharmaceutical Carriers

- Pharmaceutically acceptable carriers facilitate storage and administration of a vaccine to a subject. Examples of pharmaceutically acceptable carriers are described herein. Additional pharmaceutical acceptable carriers are well known in the art.

- Pharmaceutically acceptable carriers may contain different components such as a buffer, normal saline or phosphate buffered saline, sucrose, salts and polysorbate. An example of a pharmaceutically acceptable carrier is follows: 2.5-10 mM TRIS buffer, preferably about 5 mM TRIS buffer; 25-100 mM NaCl, preferably

about 75 mM NaCl; 2.5-10% sucrose, preferably about 5% sucrose; 0.01 -2 mM $MgCl_2$; and 0.001%-0.01% polysorbate 80 (plant derived). The pH is preferably from about 7.0-9.0, more preferably about 8.0. A specific example of a carrier contains 5 mM TRIS, 75 mM NaCl, 5% sucrose, 1 mM $MgCl_2$, 0.005% polysorbate 80 at pH 8.0.

C. Dosing Regimes

Suitable dosing regimens can be determined taking into account the efficacy of a particular vaccine and factors such as age, weight, sex and medical condition of a patient; the route of administration; the desired effect; and the number of doses. The efficacy of a particular vaccine depends on different factors such as the ability of a particular vaccine to produce polypeptide that is expressed and processed in a cell and presented in the context of MHC class I and II complexes.

HCV encoding nucleic acid administered to a patient can be part of different types of vectors including viral vectors such as adenovector, and DNA plasmid vaccines. In different embodiments concerning administration of a DNA plasmid, about 0.1 to 10 mg of plasmid is administered to a patient, and about 1 to 5 mg of plasmid is administered to a patient. In different embodiments concerning administration of a viral vector, preferably an adenoviral vector, about 10^5 to 10^{11} viral particles are administered to a patient, and about 10^7 to 10^{10} viral particles are administered to a patient.

Viral vector vaccines and DNA plasmid vaccines may be administered alone, or may be part of a prime and boost administration regimen. A mixed modality priming and booster inoculation involves either priming with a DNA vaccine and boosting with viral vector vaccine, or priming with a viral vector vaccine and boosting with a DNA vaccine.

Multiple priming, for example, about 2-4 or more may be used. The length of time between priming and boost may typically vary from about four months to a year, but other time frames may be used. The use of a priming regimen with a DNA vaccine may be preferred in situations where a person has a pre-existing anti-adenovirus immune response.

In an embodiment of the present invention, 1×10^7 to 1×10^{12} particles and preferably about 1×10^{10} to 1×10^{11} particles of adenovector is administered directly into muscle tissue. Following initial vaccination a boost is performed with an adenovector or DNA vaccine.

In another embodiment of the present invention initial vaccination is performed with a DNA vaccine directly into muscle tissue. Following initial vaccination a boost is performed with an adenovector or DNA vaccine.

- Agents such as interleukin-12, GM-CSF, B7-1, B7-2, IP10, Mig-1 can be coadministered to boost the immune response. The agents can be coadministered as proteins or through use of nucleic acid vectors.

D. Heterologous Prime-Boost

- Heterologous prime-boost is a mixed modality involving the use of one type of viral vector for priming and another type of viral vector for boosting. The heterologous prime-boost can involve related vectors such as vectors based on different adenovirus serotypes and more distantly related viruses such as adenovirus and poxvirus. The use of poxvirus and adenovirus vectors to protect mice against malaria is illustrated by Gilbert *et al.*, *Vaccine* 20:1039-1045, 2002.

- Different embodiments concerning priming and boosting involve the following types of vectors expressing desired antigens such as Mct-NS3-NS4A-NS4B-NS5A-NS5B: Ad5 vector followed by Ad6 vector; Ad6 vector followed by Ad5 vector; Ad5 vector followed by poxvirus vector; poxvirus vector followed by Ad5 vector; Ad6 vector followed by poxvirus vector; and poxvirus vector followed by Ad6 vector.

- The length of time between priming and boosting typically varies from about four months to a year, but other time frames may be used. The minimum time frame should be sufficient to allow for an immunological rest. In an embodiment, this rest is for a period of at least 6 months. Priming may involve multiple priming with one type of vector, such as 2-4 primings.

- Expression cassettes present in a poxvirus vector should contain a promoter either native to, or derived from, the poxvirus of interest or another poxvirus member. Different strategies for constructing and employing different types of poxvirus based vectors including those based on vaccinia virus, modified vaccinia virus, avipoxvirus, raccoon poxvirus, modified vaccinia virus Ankara, canarypoxviruses (such as ALVAC), fowlpoxviruses, cowpoxviruses, and NYVAC are well known in the art. (Moss, *Current Topics in Microbiology and Immunology* 158:25-38, 1982; Earl *et al.*, In *Current Protocols in Molecular Biology*, Ausubel *et al.* eds., New York: Greene Publishing Associates & Wiley Interscience; 1991:16.16.1-16.16.7, Child *et al.*, *Virology* 174(2):625-9, 1990; Tartaglia *et al.*,

Virology 188:217-232, 1992; U.S. Patent Nos., 4,603,112, 4,722,848, 4,769,330, 5,110,587, 5,174,993, 5,185,146, 5,266,313, 5,505,941, 5,863,542, and 5,942,235.

E. Adjuvants

- 5 HCV vaccines can be formulated with an adjuvant. Adjuvants are particularly useful for DNA plasmid vaccines. Examples of adjuvants are alum, AlPO_4 , alhydrogel, Lipid-A and derivatives or variants thereof, Freund's incomplete adjuvant, neutral liposomes, liposomes containing the vaccine and cytokines, non-ionic block copolymers, and chemokines.
- 10 Non-ionic block polymers containing polyoxyethylene (POE) and polyoxylpropylene (POP), such as POE-POP-POE block copolymers may be used as an adjuvant. (Newman *et al.*, *Critical Reviews in Therapeutic Drug Carrier Systems* 15:89-142, 1998.) The immune response of a nucleic acid can be enhanced using a non-ionic block copolymer combined with an anionic surfactant.
- 15 A specific example of an adjuvant formulation is one containing CRL-1005 (CytRx Research Laboratories), DNA, and benzylalkonium chloride (BAK). The formulation can be prepared by adding pure polymer to a cold ($< 5^\circ\text{C}$) solution of plasmid DNA in PBS using a positive displacement pipette. The solution is then vortexed to solubilize the polymer. After complete solubilization of the polymer a
- 20 clear solution is obtained at temperatures below the cloud point of the polymer ($\sim 6-7^\circ\text{C}$). Approximately 4 mM BAK is then added to the DNA/CRL-1005 solution in PBS, by slow addition of a dilute solution of BAK dissolved in PBS. The initial DNA concentration is approximately 6 mg/mL before the addition of polymer and BAK, and the final DNA concentration is about 5 mg/mL. After BAK addition the
- 25 formulation is vortexed extensively, while the temperature is allowed to increase from $\sim 2^\circ\text{C}$ to above the cloud point. The formulation is then placed on ice to decrease the temperature below the cloud point. Then, the formulation is vortexed while the temperature is allowed to increase from $\sim 2^\circ\text{C}$ to above the cloud point. Cooling and mixing while the temperature is allowed to increase from $\sim 2^\circ\text{C}$ to above the cloud
- 30 point is repeated several times, until the particle size of the formulation is about 200-500 nm, as measured by dynamic light scattering. The formulation is then stored on ice until the solution is clear, then placed in storage at -70°C . Before use, the formulation is allowed to thaw at room temperature.

F. Vaccine Storage

Adenovector and DNA vaccines can be stored using different types of buffers. For example, buffer A105 described in Example 9 *infra*. can be used to for vector storage.

- 5 Storage of DNA can be enhanced by removal or chelation of trace metal ions. Reagents such as succinic or malic acid, and chelators can be used to enhance DNA vaccine stability. Examples of chelators include multiple phosphate ligands and EDTA. The inclusion of non-reducing free radical scavengers, such as ethanol or glycerol, can also be useful to prevent damage of DNA plasmid from free
- 10 radical production. Furthermore, the buffer type, pH, salt concentration, light exposure, as well as the type of sterilization process used to prepare the vials, may be controlled in the formulation to optimize the stability of the DNA vaccine.

VII. EXAMPLES

- 15 Examples are provided below to further illustrate different features of the present invention. The examples also illustrate useful methodology for practicing the invention. These examples do not limit the claimed invention.

Example 1: Met-NS3-NS4A-NS4B-NS5A-NS5B Expression Cassettes

- 20 Different gene expression cassettes encoding HCV NS3-NS4A-NS4B-NS5A-NS5B were constructed based on a 1b subtype HCV BK strain. The encoded sequences had either (1) an active NS5B sequence ("NS"), (2) an inactive NS5B sequence ("NSmut"), (3) a codon optimized sequence with an inactive NS5B sequence ("NSOPTmut"). The expression cassettes also contained a CMV
- 25 promoter/enhancer and the BGH polyadenylation signal.

- The NS nucleotide sequence (SEQ. ID. NO. 5) differs from HCV BK strain GenBank accession number M58335 by 30 out of 5952 nucleotides. The NS amino acid sequence (SEQ. ID. NO. 6) differs from the corresponding 1b genotype HCV BK strain by 7 out of 1984 amino acids. To allow for initiation of translation an
- 30 ATG codon is present at the 5' end of the NS sequence. A TGA termination sequence is present at the 3' end of the NS sequence.

- The NSmut nucleotide sequence (SEQ. ID. NO. 2, Figure 2), is similar to the NS sequence. The differences between NSmut and NS include NSmut having an altered NS5B catalytic site; an optimal ribosome binding site at the 5' end; and a
- 35 TAAA termination sequence at the 3' end. The alterations in NS5B comprise bases

5138 to 5146, which encode amino acids 1711 to 1713. The alterations result in a change of amino acids GlyAspAsp into AlaAlaGly and creates an inactive form of the NS5B RNA-dependent RNA-polymerase NS5B.

- The NSOPTmut sequence (SEQ. ID. NO. 3, Figure 3) was designed based on the amino acid sequence encoded by NSmut. The NSmut amino acid sequence was back translated into a nucleotide sequence with the GCG (Wisconsin Package version 10, Genetics Computer Group, GCG, Madison, Wisc.) BACKTRANSLATE program. To generate a NSOPTmut nucleotide sequence where each amino acid is coded for by the corresponding most frequently observed human codon, the program was run choosing as parameter the generation of the most probable nucleotide sequence and specifying the codon frequency table of highly expressed human genes (human_high.cod) available within the GCG Package as translation scheme.

- Example 2: Generation pV1Jns plasmid with NS, NSmut or NSOPTmut Sequences
pV1Jns plasmids containing either the NS sequence, NSmut sequence or NSOPTmut sequences were generated and characterised as follows:

pV1Jns Plasmid with the NS Sequence

- The coding region Met-NS3-NS4A-NS4B-NS5A and the coding region Met-NS3-NS4A-NS4B-NS5A-NS5B from a HCV BK type strain (Tomei *et al.*, *J. Virol.* 67:4017-4026, 1993) were cloned into pcDNA3 plasmid (Invitrogen), generating pcD3-5a and pcD3-5b vectors, respectively. PcD3-5A was digested with Hind III, blunt-ended with Klenow fill-in and subsequently digested with Xba I, to generate a fragment corresponding to the coding region of Met-NS3-NS4A-NS4B-NS5A. The fragment was cloned into pV1Jns-poly, digested with Bgl II blunt-ended with Klenow fill-in and subsequently digested with Xba I, generating pV1JnsNS3-5A.
- pV1Jns-poly is a derivative of pV1JnsA plasmid (Montgomery *et al.*, *DNA and Cell Biol.* 12:777-783, 1993), modified by insertion of a polylinker containing recognition sites for XbaI, PmeI, PacI into the unique BglIII and NotI restriction sites. The pV1Jns plasmid with the NS sequence (pV1JnsNS3-5B) was obtained by homologous recombination into the bacterial strain BJ5183, co-transforming pV1JNS3-5A linearized with XbaI and NotI digestion and a PCR fragment containing approximately 200 bp of NS5A, NS5B coding sequence and

approximately 60 bp of the BGH polyadenylation signal. The resulting plasmid represents pV1Jns-NS.

pV1Jns-NS can be summarized as follows:

- | | | |
|----|---------------|--|
| | Bases | 1 to 1881 of pV1JnsA |
| 5 | an additional | AGCTT |
| | then the | Met-NS3-NS5B sequence (SEQ. ID. NO. 5) |
| | then the | wt TGA stop |
| | an additional | TCTAGAGCGTTTAAACCCTTAATTAAGG (SEQ. ID. NO. 14) |
| 10 | Bases | 1912 to 4909 of pV1JnsA |

pV1Jns Plasmid with the NSmut Sequence

- The V1JnsNS3-5A plasmid was modified at the 5' of the NS3 coding sequence by addition of a full Kozak sequence. The plasmid (V1JNS3-5Akozak) was obtained by homologous recombination into the bacterial strain BJ5183, co-transforming V1JNS3-5A linearized by *Afl*III digestion and a PCR fragment containing the proximal part of Intron A, the restriction site *Bgl*II, a full Kozak translation initiation sequence and part of the NS3 coding sequence.

- The resulting plasmid (V1JNS3-5Akozak) was linearized with *Xba*I digestion and co-transformed into the bacterial strain BJ5183 with a PCR fragment, containing approximately 200 bp of NS5A, the NS5B mutated sequence, the strong translation termination TAAA and approximately 60 bp of the BGH polyadenylation signal. The PCR fragment was obtained by assembling two 22bp-overlapping fragments where mutations were introduced by the oligonucleotides used for their amplification. The resulting plasmid represents pV1Jns-NSmut.

pV1Jns-NSmut can be summarized as follows:

- | | | |
|----|---------------|--|
| | Bases | 1 to 1882 of pV1JnsA |
| | then the | kozak Met-NS3-NS5B(mut) TAAA sequence (SEQ. ID. NO. 2) |
| | an additional | TCTAGA |
| 30 | Bases | 1925 to 4909 of pV1JnsA |

pV1Jns Plasmid with the NSOPTmut Sequence

- The human codon-optimized synthetic gene (NSOPTmut) with mutated NS5B to abrogate enzymatic activity, full Kozak translation initiation sequence and a strong translation termination was digested with *Bam*HI and *Sall*

restriction sites present at the 5' and 3' end of the gene. The gene was then cloned into the BglII and SalI restriction sites present in the polylinker of pV1JnsA plasmid, generating pV1Jns-NSOPTmut.

pV1Jns-NSOPTmut can be summarized as follows:

- 5 Bases 1 to 1881 of pV1JnsA
 an additional C
 then kozak Met-NS3-NS5B(optmut) TAAA sequence (SEQ. ID. NO. 3)
 an additional TTAAATGTTTAAAC (SEQ. ID. NO. 15)
 Bases 1905 to 4909 of pV1JnsA

10

Plasmids Characterization

Expression of HCV NS proteins was tested by transfection of HEK 293 cells, grown in 10% FCS/DMEM supplemented by L-glutamine (final 4 mM). Twenty-four hours before transfection, cells were plated in 6-well 35 mm diameter, to reach 90-95% confluence on the day of transfection. Forty nanograms of plasmid DNA (previously assessed as a non-saturating DNA amount) were co-transfected with 100 ng of pRSV-Luc plasmid containing the luciferase reporter gene under the control of Rous sarcoma virus promoter, using the LIPOFECTAMINE 2000 reagent. Cells were kept in a CO₂ incubator for 48 hours at 37 °C.

20

Cell extracts were prepared in 1% Triton/TEN buffer. The extracts were normalized for Luciferase activity, and run in serial dilution on 10% SDS-acrylamide gel. Proteins were transferred on nitrocellulose and assayed with antibodies directed against NS3, NS5A and NS5B to assess strength of expression and correct proteolytic cleavage. Mock-transfected cells were used as a negative control.

25

Results from representative experiments testing pV1JnsNS, pV1JnsNSmut and pV1JnsNSOPTmut are shown in Figure 12.

Example 3: Mice Immunization with Plasmid DNA Vectors

- 30 The DNA plasmids pV1Jns-NS, pV1Jns-NSmut and pV1Jns-NSOPTmut were injected in different mice strains to evaluate their potential to elicit anti-HCV immune responses. Two different strains (Balb/C and C57Black6, N=9-10) were injected intramuscularly with 25 or 50 µg of DNA followed by electrical pluses. Each animal received two doses at three weeks interval.

- 35 Humoral immune response elicited in C57Black6 mice against the NS3 protein was measured in post dose two sera by ELISA on bacterially expressed NS3

protease domain. Antibodies specific for the tested antigen were detected in animals immunized with all three vectors with geometric mean titers (GMT) ranging from 94000 to 133000 (Tables 1-3).

5

Table 1: pV1jns-NS

										GMT
Mice n.	1	2	3	4	5	6	7	8	9	
Titer	105466	891980	78799	39496	543542	182139	32351	95028	67800	94553

Table 2: pV1jns-NSmut

10

											GMT
Mice n.	11	12	13	14	15	16	17	18	19	20	
Titer	202981	55670	130786	49748	17672	174958	44304	37337	78182	193695	75083

Table 3: pV1jns-NSOPTmut

											GMT
Mice n.	21	22	23	24	25	26	27	28	29	30	
Titer	310349	43645	63496	82174	630778	297259	66861	146735	173506	77732	133165

15

A T cell response was measured in C57Black6 mice immunized with two intramuscular injections at three weeks interval with 25 μ g of plasmid DNA. Quantitative ELISpot assay was performed to determine the number of IFN γ secreting T cells in response to five pools of 20mer peptides overlapping by ten residues encompassing the NS3-NS5B sequence. Specific CD8+ response was analyzed by the same assay using a 20mer peptide encompassing a CD8+ epitope for C57Black6 mice (pep1480).

20

Cells secreting IFN γ in an antigen specific-manner were detected using a standard ELISpot assay. T cell response in C57Black6 mice immunized with two intramuscular injections at three weeks interval with 50 μ g of plasmid DNA, was

25

analyzed by the same ELISpot assay measuring the number of IFN γ secreting T cells in response to five pools of 20mer peptides overlapping by ten residues encompassing the NS3-NS5B sequence.

- Spleen cells were prepared from immunized mice and re-suspended in
 5 R10 medium (RPMI 1640 supplemented with 10% FCS, 2 mM L-Glutamine, 50 U/ml-50 μ g/ml Penicillin/Streptomycin, 10 mM Hepes, 50 μ M 2-mercapto-ethanol). Multiscreen 96-well Filtration Plates (Millipore, Cat. No. MAIPS4510, Millipore Corporation, 80 Ashby Road Bedford, MA) were coated with purified rat anti-mouse IFN γ antibody (PharMingen, Cat. No. 18181D, PharmMingen, 10975 Torreyana
 10 Road, San Diego, California 92121-1111 USA). After overnight incubation, plates were washed with PBS 1X/0.005% Tween and blocked with 250 μ l/well of R10 medium.

- Splenocytes from immunized mice were prepared and incubated for twenty-four hours in the presence or absence of 10 μ M peptide at a density of 2.5 X
 15 10⁵/well or 5 X 10⁵/well. After extensive washing (PBS 1X/0.005% Tween), biotinylated rat anti-mouse IFN γ antibody (PharMingen, Cat. No. 18112D, PharMingen, 10975 Torreyana Road, San Diego, California 92121-1111 USA) was added and incubated overnight at 4° C. For development, streptavidin-AKP (PharMingen, Cat. No. 13043E, PharMingen, 10975 Torreyana Road, San Diego,
 20 California 92121-1111 USA) and 1-StepTM NBT-BCIP development solution (Pierce, Cat. No. 34042, Pierce, P.O. Box 117, Rockford, IL 61105 USA) were added.

- Pools of 20mer overlapping peptides encompassing the entire sequence of the HCV BK strain NS3 to NS5B were used to reveal HCV-specific IFN γ -secreting T cells. Similarly a single 20mer peptide encompassing a CD8+ epitope for
 25 C57Black6 mice was used to detect CD8 response. Representative data from groups of C57Black6 and Balb/C mice (N=9-10) immunized with two injections of 25 or 50 μ g of plasmid vectors pVIJns-NS, pVIJns-NSmut and pVIJns-NSOPTmut are shown in Figures 13A and 13B.

30 Example 4: Immunization of Rhesus Macaques

Rhesus macaques (N=3) were immunized by intramuscular injection with 5mg of plasmid pVIJns-NSOPTmut in 7.5mg/ml CRL1005, Benzalkonium chloride 0.6 mM. Each animal received two doses in the deltoid muscle at 0, and 4 weeks.

CMI was measured at different time points by IFN- γ ELISPOT. This assay measures HCV antigen-specific CD8+ and CD4+ T lymphocyte responses, and can be used for a variety of mammals, such as humans, rhesus monkeys, mice, and rats.

- 5 The use of a specific peptide or a pool of peptides can simplify antigen presentation in CTL cytotoxicity assays, interferon-gamma ELISPOT assays and interferon-gamma intracellular staining assays. Peptides based on the amino acid sequence of various HCV proteins (core, E2, NS3, NS4A, NS4B, NS5A, NS5B) were prepared for use in these assays to measure immune responses in HCV DNA and
- 10 The individual peptides are overlapping 20-mers, offset by 10 amino acids. Large pools of peptides can be used to detect an overall response to HCV proteins while smaller pools and individual peptides may be used to define the epitope specificity of a response.

15

IFN γ ELISPOT

- The IFN γ -ELISPOT assay provides a quantitative determination of HCV-specific T lymphocyte responses. PBMC are serially diluted and placed in microplate wells coated with anti-rhesus IFN- γ antibody (MD-1 U-Cytech). They are
- 20 cultured with a HCV peptide pool for 20 hours, resulting in the restimulation of the precursor cells and secretion of IFN- γ . The cells are washed away, leaving the secreted IFN bound to the antibody-coated wells in concentrated areas where the cells were sitting. The captured IFN is detected with biotinylated anti-rhesus IFN antibody (detector Ab U-Cytech) followed by alkaline phosphatase-conjugated streptavidin
- 25 (Pharmingen 13043E). The addition of insoluble alkaline phosphatase substrate results in dark spots in the wells at the sites where the cells were located, leaving one spot for each T cell that secreted IFN- γ .

- The number of spots per well is directly related to the precursor frequency of antigen-specific T cells. Gamma interferon was selected as the cytokine
- 30 visualized in this assay (using species specific anti-gamma interferon monoclonal antibodies) because it is the most common, and one of the most abundant cytokines synthesized and secreted by activated T lymphocytes. For this assay, the number of spot forming cells (SFC) per million PBMCs is determined for samples in the

presence and absence (media control) of peptide antigens. Data from Rhesus macaques on PBMC from post dose two material are shown in Table 4.

Table 4

Pep pools	PV1J-NSOPTmut		
	21G	99C161	99C166
F (NS3p)	8	10	170
G (NS3h)	7	592	229
H (NS4)	3	14	16
I (NS5a)	5	71	36
L (NS5b)	14	23	11
M (NS5b)	3	35	8
DMSO	2	4	5

- 5 INFyELISPOT on PBMC from Rhesus monkeys immunized with two injections of 5 mg DNA/dose in OPTIVAX/BAK of plasmid pV1Jns-NSOPTmut. Data are expressed as SFC7 10⁶ PBMC.

Example 5: Construction of Ad6 Pre-Adenovirus Plasmids

Ad6 pre-adenovirus plasmids were obtained as follows:

10

Construction of pAd6 E1-E3+ Pre-adenovirus Plasmid

- 15 An Ad6 based pre-adenovirus plasmid which can be used to generate first generation Ad6 vectors was constructed either taking advantage of the extensive sequence identity (approx. 98%) between Ad5 and Ad6 or containing only Ad6 regions. Homologous recombination was used to clone wtAd6 sequences into a bacterial plasmid.

- 20 A general strategy used to recover pAd6E1-E3+ as a bacterial plasmid containing Ad5 and Ad6 regions is illustrated in Figure 10. Cotransformation of BJ 5183 bacteria with purified wt Ad6 viral DNA and a second DNA fragment termed the Ad5 ITR cassette resulted in the circularization of the viral genome by homologous recombination. The ITR cassette contains sequences from the right (bp 33798 to 35935) and left (bp 1 to 341 and bp 3525 to 5767) end of the Ad5 genome separated by plasmid sequences containing a bacterial origin of replication and an ampicillin resistance gene. The ITR cassette contains a deletion of E1 sequences from

Ad5 342 to 3524. The Ad5 sequences in the ITR cassette provide regions of homology with the purified Ad6 viral DNA in which recombination can occur.

Potential clones were screened by restriction analysis and one clone was selected as pAd6E1-E3+. This clone was then sequenced in its entirety. pAd6E1-E3+ contains Ad5 sequences from bp 1 to 341 and from bp 3525 to 5548, Ad6 bp 5542 to 33784, and Ad5 bp 33967 to 35935 (bp numbers refer to the wt sequence for both Ad5 and Ad6). pAd6E1-E3+ contains the coding sequences for all Ad6 virion structural proteins which constitute its serotype specificity.

A general strategy used to recover pAd6E1-E3+ as a bacterial plasmid containing Ad6 regions is illustrated in Figure 11. Cotransformation of BJ 5183 bacteria with purified wt Ad6 viral DNA and a second DNA fragment termed the Ad6 ITR cassette resulted in the circularization of the viral genome by homologous recombination. The ITR cassette contains sequences from the right (bp 35460 to 35759) and left (bp 1 to 450 and bp 3508 to 3807) end of the Ad6 genome separated by plasmid sequences containing a bacterial origin of replication and an ampicillin resistance gene. These three segments were generated by PCR and cloned sequentially into pNEB193, generating pNEBAd6-3 (the ITR cassette). The ITR cassette contains a deletion of E1 sequences from Ad5 451 to 3507. The Ad6 sequences in the ITR cassette provide regions of homology with the purified Ad6 viral DNA in which recombination can occur.

Construction of pAd6 E1-E3- pre-adenovirus plasmids

Ad6 based vectors containing A5 regions and deleted in the E3 region were constructed starting with pAd6E1-E3+ containing Ad5 regions. A 5322 bp subfragment of pAd6E1-E3+ containing the E3 region (Ad6 bp 25871 to 31192) was subcloned into pABS.3 generating pABSA6E3. Three E3 deletions were then made in this plasmid generating three new plasmids pABSA6E3(1.8Kb) (deleted for Ad6 bp 28602 to 30440), pABSA6E3(2.3Kb) (deleted for Ad6 bp 28157 to 30437) and pABSA6E3(2.6Kb) (deleted for Ad6 bp 28157 to 30788). Bacterial recombination was then used to substitute the three E3 deletions back into pAd6E1-E3+ generating the Ad6 genome plasmids pAd6E1-E3-1.8Kb, pAd6E1-E3-2.3Kb and pAd6E1-E3-2.6Kb.

Example 6: Generation of Ad5 Genome Plasmid with the NS Sequence

A pcDNA3 plasmid (Invitrogen) containing the coding region NS3-NS4A-NS4B-NS5A was digested with *XmnI* and *NruI* restriction sites and the DNA fragment containing the CMV promoter, the NS3-NS4A-NS4B-NS5A coding sequence and the Bovine Growth Hormone (BGH) polyadenylation signal was cloned into the unique *EcorV* restriction site of the shuttle vector pDelE1Spa, generating the Sva3-5A vector.

A pcDNA3 plasmid containing the coding region NS3-NS4A-NS4B-NS5A-NS5B was digested with *XmnI* and *EcoRI* (partial digestion), and the DNA fragment containing part of NS5A, NS5B gene and the BGH polyadenylation signal was cloned into the Sva3-5A vector, digested *EcoRI* and *BglIII* blunted with Klenow, generating the Sva3-5B vector.

The Sva3-5B vector was finally digested *SspI* and *BsrI*1107I restriction sites and the DNA fragment containing the expression cassette (CMV promoter, NS3-NS4A-NS4B-NS5A-NS5B coding sequence and the BGH polyadenylation signal) flanked by adenovirus sequences was co-transformed with pAd5HVO (E1-, E3-) *ClaI* linearized genome plasmid into the bacterial strain BJ5183, to generate pAd5HVONS. pAd5HVO contains Ad5 bp 1 to 341, bp 3525 to 28133 and bp 30818 to 35935.

Example 7: Generation of Adenovirus Genome Plasmids with the NSmut Sequence

Adenovirus genome plasmids containing an NS-mut sequence were generated in an Ad5 or Ad6 background. The Ad6 background contained Ad5 regions at bases 1 to 450, 3511 to 5548 and 33967 to 35935.

pV1JNS3-5Akozak was digested with *BglIII* and *XbaI* restriction enzymes and the DNA fragment containing the Kozak sequence and the sequence coding NS3-NS4A-NS4B-NS5A was cloned into a *BglIII* and *XbaI* digested polypMRKpdelE1 shuttle vector. The resulting vector was designated shNS3-5Akozak.

PolypMRKpdelE1 is a derivative of RKpdelE1(Pac/pIX/pack450) + CMVmin+BGHpA(str.) modified by the insertion of a polylinker containing recognition sites for *BglII*, *PmeI*, *SwaI*, *XbaI*, *SalI*, into the unique *BglIII* restriction site present downstream the CMV promoter. MRKpdelE1(Pac/pIX/pack450) + CMVmin + BGHpA(str.) contains Ad5 sequences from bp 1 to 5792 with a deletion of E1 sequences from bp 451 to 3510. The human CMV promoter and BGH polyadenylation signal were inserted into the E1 deletion in an E1 parallel orientation with a unique *BglIII* site separating them.

The NS5B fragment, mutated to abrogate enzymatic activity and with a strong translation termination at the 3' end, was obtained by assembly PCR and inserted into the shNS3-5Akozak vector via homologous recombination, generating polypMRKpdeIE1NSmut. In polypMRKpdeIE1NSmut the NS-mut coding sequence is under the control of CMV promoter and the BGH polyadenylation signal is present downstream.

The gene expression cassette and the flanking regions which contain adenovirus sequences allowing homologous recombination were excised by digestion with *PacI* and *BstI*1107I restriction enzymes and co-transformed with either pAd5HVO (E1-,E3-) or pAd6E1-E3-2.6Kb *ClaI* linearized genome plasmids into the bacterial strain BJ5183, to generate pAd5HVONSMut and pAd6E1-,E3-NSmut, respectively.

pAd6E1-E3-2.6Kb contains Ad5 bp 1 to 341 and from bp 3525 to 5548, Ad6 bp 5542 to 28157 and from bp 30788 to 33784, and Ad5 bp 33967 to 35935 (bp numbers refer to the wt sequence for both Ad5 and Ad6). In both plasmids the viral ITR's are joined by plasmid sequences that contain the bacterial origin of replication and an ampicillin resistance gene.

Example 8: Generation of Adenovirus Genome Plasmids with the NSOPTmut

The human codon-optimized synthetic gene (NSOPTmut) provided by SEQ. ID. NO. 3 cloned into a pCRBlunt vector (Invitrogen) was digested with *BamHI* and *Sall* restriction enzymes and cloned into *BglII* and *Sall* restriction sites present in the shuttle vector polypMRKpdeIE1. The resulting clone (polypMRKpdeIE1NSOPTmut) was digested with *PacI* and *BstI*1107I restriction enzymes and co-transformed with either pAd5HVO (E1-,E3-) or pAd6E1-E3-2.6Kb *ClaI* linearized genome plasmids, into the bacterial strain BJ5183, to generate pAd5HVONSOPTmut and pAd6E1-,E3-NSOPTmut, respectively.

Example 9: Rescue and Amplification of Adenovirus Vectors

Adenovectors were rescued in Per.6 cells. Per.C6 were grown in 10% FCS / DMEM supplemented by L-glutamine (final 4mM), penicillin/streptomycin (final 100 IU/ml) and 10 mM MgCl₂. After infection, cells were kept in the same medium supplemented by 5% horse serum (HS). For viral rescue, 2.5 X 10⁶ Per.C6 were plated in 6 cm ø Petri dishes.

Twenty-four hours after plating, cells were transfected by calcium phosphate method with 10 µg of the *Pac 1* linearized adenoviral DNA. The DNA precipitate was left on the cells for 4 hours. The medium was removed and 5% HS/DMEM was added.

- 5 Cells were kept in a CO₂ incubator until a cytopathic effect was visible (1 week). Cells and supernatant were recovered and subjected to 3X freeze/thawing cycles (liquid nitrogen / water bath at 37°C). The lysate was centrifuged at 3000 rpm at -4°C for 20 minutes and the recovered supernatant (corresponding to a cell lysate containing virus passed on cells only once; P1) was used, in the amount of 1 ml/dish, to infect 80-90% confluent Per.C6 in 10 cm ø Petri dishes. The infected cells were
10 incubated until a cytopathic effect was visible, cells and supernatant recovered and the lysate prepared as described above (P2).

- P2 lysate (4 ml) were used to infect 2 X 15 cm ø Petri dishes. The lysate recovered from this infection (P3) was kept in aliquots at -80°C as a stock of
15 virus to be used as starting point for big viral preparations. In this case, 1 ml of the stock was enough to infect 2 X 15 cm ø Petri dishes and resulting lysate (P4) was used for the infection of the Petri dishes devoted to the large scale infection.

- Further amplification was obtained from the P4 lysate which was diluted in medium without FCS and used to infect 30 X 15 cm ø Petri dishes (with
20 Per.C6 80%-90% confluent) in the amount of 10 ml/dish. Cells were incubated 1 hour in the CO₂ incubator, mixing gently every 20 minutes. 12 ml / dish of 5% HS / DMEM was added and cells were incubated until a cytopathic effect was visible (about 48 hours).

- Cells and supernatant were collected and centrifuged at 2K rpm for 20
25 minutes at 4°C. The pellet was resuspended in 15 ml of 0.1 M Tris pH=8.0. Cells were lysed by 3X freeze/thawing cycles (liquid nitrogen / water bath at 37°C). 150 µl of 2 M MgCl₂ and 75 µl of DNase (10 mg of bovine pancreatic deoxyribonuclease I in 10 ml of 20 mM Tris-HCl pH= 7.4, 50 mM NaCl, 1 mM dithiothreitol, 0.1 mg/ml bovine serum albumin, 50% glycerol) were added. After a 1 hour incubation at 37°C
30 in a water bath (vortex every 15 minutes) the lysate was centrifuged at 4K rpm for 15 minutes at 4°C. The recovered supernatant was ready to be applied on CsCl gradient.

The CsCl gradients were prepared in SW40 ultra-clear tubes as

follows:

- 0.5 ml of 1.5d CsCl
35 3 ml of 1.35d CsCl

3 ml of 1.25d CsCl

5-ml/ tube of viral supernatant was applied.

If necessary, the tubes were topped up with 0.1 M tris-Cl pH=8.0.

Tubes were centrifuged at 35K rpm for 1 hour at -10°C with rotor SW40. The viral bands (located at the 1.25/1.35 interface) were collected using a syringe.

The virus was transferred into a new SW40 ultraclear tube and 1.35d CsCl was added to top the tube up. After centrifugation at 35K rpm for 24 hours at 10°C in the rotor SW40, the virus was collected in the smallest possible volume and dialyzed extensively against buffer A105 (5 mM Tris, 5% sucrose, 75 mM NaCl, 1 mM MgCl_2 , 0.005% polysorbate 80 pH=8.0). After dialysis, glycerol was added to final 10% and the virus was stored in aliquots at -80°C .

Example 10: Enhanced Adenovector Rescue

First generation Ad5 and Ad6 vectors carrying HCV NSOPTmut transgene were found to be difficult to rescue. A possible block in the rescue process might be attributed to an inefficient replication of plasmid DNA that is a sub-optimal template for the replication machinery of adenovirus. The absence of the terminal protein linked to the 5' ends of the DNA (normally present in the viral DNA), associated with the very high G-C content of the transgene inserted in the E1 region of the vector, may be causing a substantial reduction in replication rate of the plasmid-derived adenovirus.

To set up a more efficient and reproducible procedure for rescuing Ad vectors, an expression vector (pE2; Figure 19) containing all E2 proteins (polymerase, pre-terminal protein and DNA binding protein) as well as E4 orf6 under the control of tet-inducible promoter was employed. The transfection of pE2 in combination with a normal preadeno plasmid in PerC6 and in 293 leads to a strong increase of Ad DNA replication and to a more efficient production of complete infectious adenovirus particles.

Plasmid Construction

pE2 is based on the cloning vector pBI (CLONTECH) with the addition of two elements to allow episomal replication and selection in cell culture: (1) the EBV-oriP (EBV [nt] 7421-8042) region permitting plasmid replication in synchrony with the cell cycle when EBNA-1 is expressed and (2) the hygromycin-B phosphotransferase (HPH)-resistance gene allowing a positive selection of

transformed cells. The two transcriptional units for the adenoviral genes E2 a and b and E4-Orf6 were constructed and assembled in pE2 as described below.

The Ad5-Polymerase *Clal/SphI* fragment and the Ad5-pTP *Acc65/EcoRV* fragment were obtained from pVac-Pol and pVac-pTP (Stunnenberg *et al. NAR* 16:2431-2444, 1988). Both fragments were filled with Klenow and cloned into the *Sall* (filled) and *EcoRV* sites of pBI, respectively obtaining pBI-Pol/pTP.

EBV-OriP element from pCEP4 (Invitrogen) was first inserted within two chicken β -globin insulator dimers by cloning it into *BamHI* site of pJC13-1 (Chung *et al., Cell* 74(3):505-14, 1993). HS4-OriP fragment from pJC13-OriP was then cloned inside pSA1mv (a plasmid containing tk-Hygro-B resistance gene expression cassette as well as Ad5 replication origin), the ITR's arranged as head-to-tail junction, obtained by PCR from pFG140 (Graham, *EMBO J.* 3:2917-2922, 1984) using the following primers: 5'-TCGAATCGATACGCGAACCTACGC-3' (SEQ. ID. NO. 16) and 5'-TCGACGTGTCGACTTCGAAGCGCACCAAAAACGTC-3' (SEQ. ID. NO. 17), thus generating pMVHS4OriP. A DNA fragment from pMVHS4OriP, containing the insulated OriP, Ad5 ITR junction and tk-HygroB cassette, was then inserted into pBI-Pol/pTP vector restricted *AseI/AatII* generating pBI-Pol/pTPHS4.

To construct the second transcriptional unit expressing Ad5-Orf6 as well as Ad5-DBP, E4orf6 (Ad 5 [nt] 33193-34077) obtained by PCR was first inserted into pBI vector, generating pBI-Orf6. Subsequently, DBP coding DNA sequence (Ad 5 [nt] 22443-24032) was inserted into pBI-Orf6 obtaining the second bi-directional Tet-regulated expression vector (pBI-DBP/E4orf6). The original polyA signals present in pBI were substituted with BGH and SV40 polyA.

pBI-DBP/E4orf6 was then modified by inserting a DNA fragment containing the Adeno5-ITRs arranged in head-to-tail junction plus the hygromycin B resistance gene obtained from plasmid pSA-1mv. The new plasmid pBI-DBP/E4orf6shuttle was then used as donor plasmid to insert the second tet-regulated transcriptional unit into pBI-Pol/pTPHS4 by homologous recombination using *E. coli* strain BJ5183 obtaining pE2.

Cell lines, Transfections and Virus Amplification

PerC6 cells were cultured in Dulbecco's modified Eagle's Medium (DMEM) plus 10% fetal bovine serum (FBS), 10 mM $MgCl_2$, penicillin (100 U/ml), streptomycin (100 μ g/ml) and 2 mM glutamine.

All transient transfections were performed using Lipofectamine2000 (Invitrogen) as described by the manufacturer. 90% confluent PERC.6™ planted in 6-cm plates were transfected with 3.5 µg of Ad5/6NSOPTmut pre-adeno plasmids, digested with PacI, alone or in combination with 5 µg pE2 plus 1 µg pUHD52.1. 5 pUHD52.1 is the expression vector for the reverse tet transactivator 2 (rtTA2) (Urlinger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 97(14):7963-7968, 2000). Upon transfection, cells were cultivated in the presence of 1 µg/ml of doxycycline to activate pE2 expression. 7 days post-transfection cells were harvested and cell lysate was obtained by three cycles of freeze-thaw. Two ml of cell lysate were used to infect 10 a second 6-cm dish of PerC6. Infected cells were cultivated until a full CPE was observed then harvested. The virus was serially passaged five times as described above, then purified on CsCl gradient. The DNA structure of the purified virus was controlled by endonuclease digestion and agarose gel electrophoresis analysis and compared to the original pre-adeno plasmid restriction pattern.

15

Example 11: Partial Optimization of HCV Polyprotein Encoding Nucleic acid

Partial optimization of HCV polyprotein encoding nucleic acid was performed to facilitate the production of adenovectors containing codons optimized for expression in a human host. The overall objective was to provide for increased 20 expression due to codon optimization, while facilitating the production of an adenovector encoding HCV polyprotein.

Several difficulties were encountered in producing an adenovector encoding HCV polyprotein with codons optimized for expression in a human host. An adenovector containing an optimized sequence (SEQ. ID. NO. 3) was found to be 25 more difficult to synthesize and rescue than an adenovector containing a non-optimized sequence (SEQ. ID. NO. 2).

The difficulties in producing an adenovector containing SEQ. ID. NO. 3 were attributed to a high GC content. A particularly problematic region was the region at about position 3900 of NSOPTmut (SEQ. ID. NO. 3).

30 Alternative versions of optimized HCV encoding nucleic acid sequence were designed to facilitate its use in an adenovector. The alternative versions, compared to NSOPTmut, were designed to have a lower overall GC content, to reduce/avoid the presence of potentially problematic motifs of consecutive G's or C's, while maintaining a high level of codon optimization to allow improved 35 expression of the encoded polyprotein and the individual cleavage products.

A starting point for the generation of a suboptimally codon-optimized sequence is the coding region of the NSOPTmut nucleotide sequence (bases 7 to 5961 of SEQ. ID. NO. 3). Values for codon usage frequencies (normalized to a total of 1.0 for each amino acid) were taken from the file human_high.cod available in the Wisconsin Package Version 10.3 (Accelrys Inc., a wholly owned subsidiary of Pharmacoceia, Inc).

To reduce the local and overall GC content a table defining preferred codon substitutions for each amino acid was manually generated. For each amino acid the codon having 1) a lower GC content as compared to the most frequent codon and 2) a relatively high observed codon usage frequency (as defined in human_high.cod) was chosen as the replacement codon. For example for Arg the codon with the highest frequency is CGC. Out of the other five alternative codons encoding Arg (CGG, AGG, AGA, CGT, CGA) three (AGG, CGT, CGA) reduce the GC content by 1 base, one (AGA) by two bases and one (CGG) by 0 bases. Since the AGA codon is listed in human_high.cod as having a relatively low usage frequency (0.1), the codon substituting CGC was therefore chosen to be AGG with a relative frequency of 0.18. Similar criteria were applied in order to establish codon replacements for the other amino acids resulting in the list shown in Table 5. Parameters applied in the following optimization procedure were determined empirically such that the resulting sequence maintained a considerably improved codon usage (for each amino acid) and the GC content (overall and in form of local stretches of consecutive G's and/or C's) was decreased.

Two examples of partial optimized HCV encoding sequences are provided by SEQ. ID. NO. 10 and SEQ. ID. NO. 11. SEQ. ID. NO. 10 provides a HCV encoding sequence that is partially optimized throughout. SEQ. ID. NO. 11 provides an HCV encoding sequence fully optimized for codon usage with the exception of a region that was partially optimized.

Codon optimization was performed using the following procedure:

Step 1) The coding region of the input fully optimized NSOPTmut sequence was analyzed using a sliding window of 3 codons (9 bases) shifting the window by one codon after each cycle. Whenever a stretch containing 5 or more consecutive C's and/or G's was detected in the window the following replacement rule was applied: Let N indicate the number of codon replacements previously performed. If N is odd replace the middle codon in the window with the codon specified in Table 5, if N is even replace the third terminal codon in the window with the codon

specified in a codon optimization table such as human_high.cod. If Leu or Val is present at the second or third codon do not apply any replacement in order not to introduce Leu or Val codons with very low relative codon usage frequency (see, for example, human_high.cod). In the following cycle analysis of the shifted window was then applied to a sequence containing the replacements of the previous cycle.

The alternating replacement of the middle and terminal codon in the 3 codon window was found empirically to give a more satisfying overall maintenance of optimized codon usage while also reducing GC content (as judged from the final sequence after the procedure). In general, however, the precise replacement strategy depends on the amino acid sequence encoded by the nucleotide sequence under analysis and will have to be determined empirically.

Step 2) The sequence containing all the codon replacements performed during step 1) was then subjected to an additional analysis using a sliding window of 21 codons (63 bases) in length: according to an adjustable parameter the overall GC content in the window was determined. If the GC content in the window was higher than 70% the following codon replacement strategy was applied: In the window replace the codons for the amino acids Asn, Asp, Cys, Glu, His, Ile, Lys, Phe, Tyr by the codons given in Table 5. Restriction of the replacement to this set of amino acids was motivated by the fact that a) the replacement codon still has an acceptably high frequency of usage in human_high.cod and b) the average overall human codon usage in CUTG for the replacement codon is nearly as high as the most frequent codon. In the following cycle analysis of the shifted window is then applied to a sequence containing the replacements of the previous cycle.

The threshold 70% was determined empirically by compromising between an overall reduction in GC content and maintenance of a high codon optimization for the individual amino acids. As in step 1) the precise replacement strategy (choice of amino acids and GC content threshold value) will again depend on the amino acid sequence encoded by the nucleotide sequence under analysis and will have to be determined empirically.

Step 3) The sequence generated by steps 1) and 2) was then manually edited and additional codons were changed according to the following criteria: Regions still having a GC content higher than 70% over a window of 21 codons were examined manually and a few codons were replaced again following the scheme given in Table 5.

Subsequent steps were performed to provide for useful restriction sites, remove possible open reading frames on the complementary strand, to add homologous recombinant regions, to add a Kozak signal, and to add a terminator. These steps are numbered 4-7

5 Step 4) The sequence generated in step 3 was examined for the absence of certain restriction sites (BglII, PmeI and XbaI) and presence of only 1 StuI site to allow a subsequent cloning strategy using a subset of restriction enzymes. Two sites (one for BglII and one for StuI) were removed from the sequence by replacing codons that were part of the respective recognition sites.

10 Step 5) The sequence generated by steps 1) through 4) was then modified according to allow subsequent generation of a modified NSOPTmut sequence (by homologous recombination). In the sequence obtained from steps 1) through 4) the segment comprising base 3556 to 3755 and the segment comprising base 4456 to 4656 were replaced by the corresponding segments from NSOPTmut.
15 The segment comprising bases 3556 to 4656 of SEQ. ID. NO. 10 can be used to replace the problematic region in NSOPTmut (around position 3900) by homologous recombination thus creating the variant of NSOPTmut having the sequence of SEQ. ID. NO. 11.

20 Step 6) Analysis of the sequence generated through steps 1) to 5) revealed a potential open reading frame spanning nearly the complete fragment on the complementary strand. Removal of all codons CTA and TTA (Leu) and TCA (Ser) from the sense strand effectively removed all stop codons in one of the reading frames on the complementary strand. Although the likelihood for transcription of this complementary strand open reading frame and subsequent translation into protein is
25 very small, in order to exclude a potential interference with the transcription and subsequent translation of the sequence encoded on the sense strand, TCA codons for Ser were introduced on the sense approximately every 500 bases. No changes were introduced in the segments introduced during step 5) to allow homologous recombination. The TCA codon for Ser was preferred over the CTA and TTA codons
30 for Leu because of the higher relative frequency for TCA (0.05) as compared to TAA (0.02) and TTA (0.03) in human_high.cod. In addition, the average human codon usage from CUTG favored TCA (0.14 against 0.07 for CTA and TTA).

35 Step 7) In a final step GCCACC was added at the 5' end of the sequence to generate an optimized internal ribosome entry site (Kozak signal) and a TAAA stop signal was added at the 3'. To maintain the initiation of translation

properties of NSsuboptmut the first 8 codons of the coding region were kept identical to the NSOPTmut sequence. The resulting sequence was again checked for the absence of BglII, PmeI and XbaI recognition sites and the presence of only 1 StuI site.

The NSsuboptmut sequence (SEQ. ID. NO. 10) has an overall reduced

- 5 GC content (63.5%) as compared to NSOPTmut (70.3%) and maintains a well optimized level of codon usage optimization. Nucleotide sequence identity of NSsuboptmut is 77.2% with respect to NSmut.

Table 5: Definition of codon replacements performed during steps 1) and 2).

10

Amino Acid	Most frequent codon	Relative frequency	Reduction in GC content (bases)	Replacement codon	Relative frequency
Amino Acids where the replacement codon reduces the codon GC-content by 1 base					
Ala	GCC	0.51	1	GCT	0.17
Arg	CGC	0.37	1	AGG	0.18
Asn	AAC	0.78	1	AAT	0.22
Asp	GAC	0.75	1	GAT	0.25
Cys	TGC	0.68	1	TGT	0.32
Glu	GAG	0.75	1	GAA	0.25
Gln	CAG	0.88	1	CAA	0.12
Gly	GGC	0.50	1	GGA	0.14
His	CAC	0.79	1	CAT	0.21
Ile	ATC	0.77	1	ATT	0.18
Lys	AAG	0.82	1	AAA	0.18
Phe	TTC	0.80	1	TTT	0.20
Pro	CCC	0.48	1	CCT	0.19
Ser	AGC	0.34	1	TCT	0.13
Thr	ACC	0.51	1	ACA	0.14
Tyr	TAC	0.74	1	TAT	0.26
Amino Acids with no alternative codon					
Met	ATG	1.00	0	ATG	1.00
Trp	TGG	1.00	0	TGG	1.00

Amino Acids where the replacement codon has a very low relative frequency. These amino acids were excluded from the replacement procedure					
Leu	CTG	0.58	1	TTG	0.06
Val	GTG	0.64	1	GTT	0.07

Example 12: Virus Characterization

Adenovectors were characterized by: (a) measuring the physical particles/ml; (b) running a TaqMan PCR assay; and (c) checking protein expression after infection of HeLa cells.

a) Physical Particles Determination

CsCl purified virus was diluted 1/10 and 1/100 in 0.1% SDS PBS. As a control, buffer A105 was used. These dilutions were incubated 10 minutes at 55°C. After spinning the tubes briefly, O.D. at 260 nm was measured. The amount of viral particles was calculated as follows: 1 OD 260 nm = 1.1×10^{12} physical particles/ml. The results were typically between 5×10^{11} and 1×10^{12} physical particles /ml.

b) TaqMan PCR Assay

TaqMan PCR assay was used for adenovectors genome quantification (Q-PCR particles/ml). TaqMan PCR assay was performed using the ABI Prism 7700-sequence detector. The reaction was performed in a final 50 μ l volume in the presence of oligonucleotides (at final 200 nM) and probe (at final 200 μ M) specific for the adenoviral backbone. The virus was diluted 1/10 in 0.1% SDS PBS and incubated 10 minutes at 55°C. After spinning the tube briefly, serial 1/10 dilutions (in water) were prepared. 10 μ l the 10^{-3} , 10^{-5} and 10^{-7} dilutions were used as templates in the PCR assay.

The amount of particles present in each sample was calculated on the basis of a standard curve run in the same experiment. Typically results were between 1×10^{12} and 3×10^{12} Q-PCR particles /ml.

c) Expression of HCV Non-Structural Proteins

Expression of HCV NS proteins was tested by infection of HeLa cells. Cells were plated the day before the infection at 1.5×10^6 cells/dish (10 cm ϕ Petri dishes). Different amounts of CsCl purified virus corresponding to m.o.i. of 50, 250

and 1250 pp/cell were diluted in medium (FCS free) up to a final volume of 5 ml. The diluted virus was added on the cells and incubated for 1 hour at 37°C in a CO₂ incubator (gently mixing every 20 minutes). 5 ml of 5% HS-DMEM was added and the cells were incubated at 37°C for 48 hours.

- 5 Cell extracts were prepared in 1% Triton/TEN buffer. The extracts were run on 10% SDS-acrylamide gel, blotted on nitrocellulose and assayed with antibodies directed against NS3, NS5a and NS5b in order to check the correct polyprotein cleavage. Mock-infected cells were used as a negative control. Results from representative experiments testing the Ad5-NS, MRKAd5-NSmut, MRKAd6-NSmut and MRKAd6-NSOPTmut are shown in Figure 14.

Example 13: Mice Immunization with Adenovectors Encoding Different NS Cassettes

- The adenovectors Ad5-NS, MRKAd5-NSmut, MRKAd6-NSmut and MRKAd6-NSOPTmut were injected in C57Black6 mice strains to evaluate their potential to elicit anti-HCV immune responses. Groups of animals (N=9-10) were injected intramuscularly with 10⁹ pp of CsCl purified virus. Each animal received two doses at three weeks interval.

- Humoral immune response against the NS3 protein was measured in post dose two sera from C57Black6 immunized mice by ELISA on bacterially expressed NS3 protease domain. Antibodies specific for the tested antigen were detected with geometric mean titers (GMT) ranging from 100 to 46000 (Tables 6, 7, 8 and 9).

25 Table 6: Ad5-NS

											GMT
Mice n.	1	2	3	4	5	6	7	8	9	10	
Titer	50	253	50	50	50	2257	504	50	50	50	108

Table 7: Ad5-NSmut

											GMT
Mice n.	11	12	13	14	15	16	17	18	19	20	
Titer	3162	78850	87241	6796	12134	3340	18473	13093	76167	49593	23645

Table 8: MRKAd6-NSmut

5

											GMT
Mice n.	21	22	23	24	25	26	27	28	29	30	
Titer	125626	39751	40187	65834	60619	69933	21555	49348	29290	26859	46461

Table 9: MRKAd6-NSOPTmut

								GMT
Mice n.	31	32	33	34	35	36	37	
Titer	25430	3657	893	175	10442	49540	173	2785

- 10 T cell response in C57Black6 mice was analyzed by the quantitative ELISPOT assay measuring the number of IFN γ secreting T cells in response to five pools (named from F to L+M) of 20mer peptides overlapping by ten residues encompassing the NS3-NS5B sequence. Specific CD8+ response induced in C57Black6 mice was analyzed by the same assay using a 20mer peptide
- 15 encompassing a CD8+ epitope for C57Black6 mice (pep1480). Cells secreting IFN γ in an antigen specific-manner were detected using a standard ELISpot assay.

- Spleen cells, splenocytes and peptides were produced and treated as described in Example 3, *supra*. Representative data from groups of C57Black6 mice (N=9-10) immunized with two injections of 10^9 viral particles of vectors Ad5-NS, MRKAd5-NSmut and MRKAd6-NSmut are shown in Figure 15.
- 20

Example 14: Immunization of Rhesus macaques with Adenovectors

Rhesus macaques (N=3-4) were immunized by intramuscular injection of CsCl purified Ad5-NS, MRKAd5-NSmut, MRKAd6-NSmut or MRKAd6-

NSOPTmut virus. Each animal received two doses of 10^{11} or 10^{10} vp in the deltoid muscle at 0, and 4 weeks.

CMI was measured at different time points by a) IFN- γ ELISPOT (see Example 3, *supra*), b) IFN- γ ICS and c) bulk CTL assays. These assays measure HCV antigen-specific CD8+ and CD4+ T lymphocyte responses, and can be used for a variety of mammals, such as humans, rhesus monkeys, mice, and rats.

The use of a specific peptide or a pool of peptides can simplify antigen presentation in CTL cytotoxicity assays, interferon-gamma ELISPOT assays and interferon-gamma intracellular staining assays. Peptides based on the amino acid sequence of various HCV proteins (core, E2, NS3, NS4A, NS4B, NS5a, NS5b) were prepared for use in these assays to measure immune responses in HCV DNA and adenovirus vector vaccinated rhesus monkeys, as well as in HCV-infected humans. The individual peptides are overlapping 20-mers, offset by 10 amino acids. Large pools of peptides can be used to detect an overall response to HCV proteins while smaller pools and individual peptides may be used to define the epitope specificity of a response.

IFN- γ ICS

For IFN- γ ICS, 2×10^6 PBMC in 1 ml R10 (RPMI medium, supplemented with 10% FCS) were stimulated with peptide pool antigens. Final concentration of each peptide was 2 μ g/ml. Cells were incubated for 1 hour in a CO₂ incubator at 37°C and then Brefeldin A was added to a final concentration of 10 μ g/ml to inhibit the secretion of soluble cytokines. Cells were incubated for additional 14-16 hours at 37°C.

Stimulation was done in the presence of co-stimulatory antibodies: CD28 and CD49d (anti-humanCD28 BD340975 and anti-humanCD49d BD340976). After incubation, cells were stained with fluorochrome-conjugated antibodies for surface antigens: anti-CD3, anti-CD4, anti-CD8 (CD3-APC Biosource APS0301, CD4-PE BD345769, CD8-PerCP BD345774).

To detect intracellular cytokines, cells were treated with FACS permeabilization buffer 2 (BD340973), 2x final concentration. Once fixed and permeabilized, cells were incubated with an antibody against human IFN- γ , IFN- γ FITC (Biosource AHC4338).

Cells were resuspended in 1% formaldehyde in PBS and analyzed at FACS within 24 hours. Four color FACS analysis was performed on a FACSCalibur

instrument (Becton Dickinson) equipped with two lasers. Acquisition was done gating on the lymphocyte population in the Forward versus Side Scatter plot coupled with the CD3, CD8 positive populations. At least 30,000 events of the gate were taken. The positive cells are expressed as number of IFN- γ expressing cells over 10^6 lymphocytes.

IFN- γ ELISPOT and IFN- γ ICS data from immunized monkeys after one or two injections of 10^{10} or 10^{11} vp of the different adenovectors are reported in Figures 16A-16D, 17A, and 17B.

10 *Bulk CTL Assays*

A distinguishing effector function of T lymphocytes is the ability of subsets of this cell population to directly lyse cells exhibiting appropriate MHC-associated antigenic peptides. This cytotoxic activity is most often associated with CD8+ T lymphocytes.

15 PBMC samples were infected with recombinant vaccine viruses expressing HCV antigens *in vitro* for approximately 14 days to provide antigen restimulation and expansion of memory T cells. Cytotoxicity against autologous B cell lines treated with peptide antigen pools was tested.

The lytic function of the culture is measured as a percentage of specific
20 lysis resulted from chromium released from target cells during 4 hours incubation with CTL effector cells. Specific cytotoxicity is measured and compared to irrelevant antigen or excipient-treated B cell lines. This assay is semi-quantitative and is the preferred means for determining whether CTL responses were elicited by the vaccine. Data after two injections from monkeys immunized with 10^{11} vp/dose with
25 adenovectors Ad5-NS, MRKAd5-NSmut and MRKAd6-NSmut are reported in Figures 18A-18F.

Other embodiments are within the following claims. While several
embodiments have been shown and described, various modifications may be made
30 without departing from the spirit and scope of the present invention.

WHAT IS CLAIMED IS:

1. A nucleic acid comprising a nucleotide sequence encoding a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ ID NO: 1, provided that said polypeptide has sufficient protease activity to process itself to produce an NS5B protein and said NS5B protein is enzymatically inactive.
2. The nucleic acid of claim 1, wherein said nucleotide sequence is substantially similar to the coding sequence of SEQ ID NO: 2.
3. The nucleic acid of claim 1, wherein said nucleotide sequence encodes for the polypeptide of SEQ ID NO: 1.
4. The nucleic acid of claim 3, wherein said nucleotide sequence is the coding sequence of either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.
5. The nucleic acid of claim 3, wherein said nucleotide sequence is the coding sequence of either SEQ ID NO: 2 or SEQ ID NO: 3.
6. The nucleic acid of any one of claims 1-5, wherein said nucleic acid is an expression vector capable of expressing said polypeptide from said nucleotide sequence in a human cell.
7. A nucleic acid comprising a gene expression cassette able to express a Met-NS3-NS4A-NS4B-NS5A-NS5B polypeptide substantially similar to SEQ ID NO: 1 in a human cell, provided that said polypeptide can process itself to produce an NS5B protein and said NS5B protein is enzymatically inactive, said expression cassette comprising:
 - a) a promoter transcriptionally coupled to a nucleotide sequence encoding said polypeptide;
 - b) a 5' ribosome binding site functionally coupled to said nucleotide sequence,

c) a terminator joined to the 3' end of said nucleotide sequence, and
d) a 3' polyadenylation signal functionally coupled to said nucleotide sequence.

5 8. The nucleic acid of claim 7, wherein said nucleotide sequence is substantially similar to either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

10 9. The nucleic acid of claim 8, wherein said nucleic acid is a shuttle vector further comprising a selectable marker, an origin of replication, a first adenovirus homology region and a second adenovirus homology region flanking said expression cassette, wherein said first homology region has at least about 100 base pairs substantially homologous to at least right end of a wild-type adenovirus region from about base pairs 1-425, and said second homology region has at least about 100
15 base pairs substantially homologous to at least the left end of a wild-type adenovirus region from about base pairs 3511-5792 of Ad5 or corresponding region of another adenovirus.

20 10. The nucleic acid of claim 9, wherein said nucleotide sequence encodes for a polypeptide of SEQ ID NO: 1.

 11. The nucleic acid of claim 9, wherein said nucleotide sequence is SEQ ID NO: 2.

25 12. The nucleic acid of claim 9, wherein said nucleotide sequence is either SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

 13. The nucleic acid of claim 8, wherein said nucleic acid is a plasmid suitable for administration into a human and further comprises a prokaryotic origin of replication and a gene coding for a selectable marker.
30

 14. The nucleic acid of claim 13, wherein said nucleotide sequence encodes for a polypeptide of SEQ ID NO: 1.

15. The nucleic acid of claim 14, wherein said nucleotide sequence is the coding sequence of either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.
- 5 16. The nucleic acid of claim 14, wherein said nucleotide sequence is the coding sequence of SEQ ID NO: 2 or SEQ ID NO: 3.
17. The nucleic acid of claim 14, wherein said promoter is the human intermediate early cytomegalovirus promoter (intron A), said 5' ribosome
10 binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the bovine growth hormone (BGH) polyadenylation signal.
18. The nucleic acid of claim 8, wherein said nucleic acid is a
15 adenovirus genome plasmid comprising a selectable marker, an origin of replication, and a recombinant adenovector genome containing an E1 deletion, an E3 deletion, and said expression cassette.
19. The nucleic acid of claim 8, wherein said nucleic acid is a
20 adenovirus genome plasmid comprising a selectable marker, an origin of replication, and
- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
 - b) said gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to said first region;
 - 25 c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said expression cassette;
 - d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair
30 28156 corresponding to Ad6, joined to said second region;
 - e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and

f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region.

5 20. The nucleic acid of claim 19, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region corresponds to Ad5.

10 21. The nucleic acid of claim 20, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

15 22. The nucleic acid of claim 21, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

20 23. The nucleic acid of claim 19, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 or Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.

25 24. The nucleic acid of claim 23, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

30 25. The nucleic acid of claim 24, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

35 26. The nucleic acid of claim 24, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2 or SEQ ID NO: 3.

27. The nucleic acid of claim 8, wherein said nucleic acid is a
adenovirus genome plasmid comprising an origin of replication, a selectable marker,
and:
- 5 a) a first adenovirus region from about base pair 1 to about base
pair 450 corresponding to either Ad5 or Ad6;
- b) a second adenovirus region from about base pair 3511 to about
base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair
5541 corresponding to Ad6, joined to said first region;
- 10 c) a third adenovirus region from about base pair 5549 to about
base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair
28156 corresponding to Ad6, joined to said second region;
- d) said gene expression cassette in a E3 parallel or E3 anti-parallel
orientation joined to said third region;
- 15 e) a fourth adenovirus region from about base pair 30818 to about
base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base
pair 33784 corresponding to Ad6, joined to said gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about
base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base
20 pair 35759 corresponding to Ad6, joined to said fourth region.
28. The nucleic acid of claim 27, wherein said first region
corresponds to Ad5, said second region corresponds to Ad5, said third region
corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region
25 corresponds to Ad5.
29. The nucleic acid of claim 28, wherein said promoter is the
human intermediate early cytomegalovirus promoter, said 5' ribosome binding site
consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation
30 signal.
30. The nucleic acid of claim 27, wherein said first region
corresponds to Ad5 or Ad6, said second region corresponds to Ad5 of Ad6, said third
region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth
35 region corresponds to Ad5 or Ad6.

31. The nucleic acid of claim 30, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.
32. The nucleic acid of claim 8, wherein said nucleic acid is a adenovector consisting of a nucleotide sequence substantially similar to of SEQ ID NO. 4 or a derivative thereof, wherein said derivative thereof has the HCV polyprotein encoding sequence present in SEQ ID NO: 4 replaced with the HCV polyprotein encoding sequence of either SEQ ID NO: 3, SEQ ID NO: 10 or SEQ ID NO: 11.
33. The nucleic acid of claim 8, wherein said nucleic acid is an adenovector having an adenovector genome containing an E1 deletion, an E3 deletion, and said expression cassette
34. The nucleic acid of claim 8, wherein said nucleic acid is an adenovector consisting of:
- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
 - b) said gene expression cassette in a E1 parallel or E1 anti-parallel orientation joined to said first region;
 - c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said expression cassette;
 - d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
 - e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and
 - f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region.

35. The nucleic acid of claim 34, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region corresponds to Ad5.

36. The nucleic acid of claim 35, wherein said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

37. The nucleic acid of claim 36, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is either SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

38. The nucleic acid of claim 34, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 or Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.

39. The nucleic acid of claim 37, where said promoter is the human intermediate early cytomegalovirus promoter, said 5' ribosome binding site consists of SEQ ID NO: 12, and said 3' polyadenylation is the BGH polyadenylation signal.

40. The nucleic acid of claim 39, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 10, or SEQ ID NO: 11.

41. The nucleic acid of claim 39, wherein said expression cassette is in an E1 anti parallel orientation and said nucleotide sequence is SEQ ID NO: 2 or SEQ ID NO: 3.

42. The nucleic acid of claim 8, wherein said nucleic acid is an adenovector consisting of:

- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;
- c) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- d) said gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to said third region;
- e) a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region.

43. The nucleic acid of claim 42, wherein said first region corresponds to Ad5, said second region corresponds to Ad5, said third region corresponds to Ad5, said fourth region corresponds to Ad5, and said fifth region corresponds to Ad5.

44. The nucleic acid of claim 42, wherein said first region corresponds to Ad5 or Ad6, said second region corresponds to Ad5 or Ad6, said third region corresponds to Ad6, said fourth region corresponds to Ad6, and said fifth region corresponds to Ad5 or Ad6.

45. An adenovector consisting of the nucleic acid sequence of SEQ ID NO. 4 or a derivative thereof, wherein said derivative thereof has the HCV polypeptide encoding sequence present in SEQ ID NO: 4 replaced with the HCV polypeptide encoding sequence of either SEQ ID NO: 3, SEQ ID NO: 10 or SEQ ID NO: 11.

46. An adenovector produced by a process comprising the steps of:

- a) producing an adenovirus genome plasmid by homologous recombination between the shuttle vector of claim 9 and a nucleic acid comprising;
a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- 5 a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;
a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- 10 a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and
a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region; and
- 15 b) rescuing said adenovector from said adenovirus plasmid.
47. A cultured recombinant cell comprising the nucleic acid of claim 6.
- 20 48. A cultured recombinant cell comprising the nucleic acid of any one of claims 9-46.
- 25 49. A method of making an adenovector comprising the steps of:
a) producing an adenovirus genome plasmid comprising a gene expression cassette by homologous recombination between the nucleic acid of claim 9 and a nucleic acid comprising;
a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
- 30 a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;

a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;

- 5 a fourth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said third region; and

a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to the fourth region; and

- 10 b) rescuing said recombinant adenovirus from said recombinant adenovirus plasmid.

50. A pharmaceutical composition comprising the nucleic acid of any one of claims 13-17 and 32-46 and pharmaceutically acceptable carrier.

15

51. A method of treating a patient comprising the step of administering to said patient an effective amount of the nucleic acid of any one of claims 13-17 and 32-46.

20

52. The method of claim 51, wherein said patient is a human.

53. The method of claim 52, wherein said patient is not infected with HCV.

25

54. The method of claim 52, wherein said patient is infected with HCV.

55. A recombinant nucleic acid comprising one or more Ad6 regions and a region not present in Ad6, wherein at least one Ad6 region is selected from the group consisting of: E1A, E1B, E2B, E2A, E4, L1, L2, L4, and L5.

30

56. The recombinant nucleic acid of claim 55, wherein said region not present in Ad6, is an expression cassette coding for a polypeptide not found in Ad6.

35

57. The recombinant nucleic acid of claim 56, wherein said recombinant nucleic acid is an adenovirus vector defective in at least E1 that is able to replicate when E1 is supplied *in trans*.

- 5 58. The recombinant nucleic acid of claim 57, wherein said vector consists of:
- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;
 - b) said gene expression cassette in an E1 parallel or E1 anti-parallel orientation joined to said first region;
 - 10 c) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said gene expression cassette;
 - d) a third adenovirus region from about base pair 5549 to about base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
 - 15 e) an optionally present fourth region from about base pair 28134 to about base pair 30817 corresponding to Ad5, or from about base pair 28157 to about 30789 corresponding to Ad6, joined to said third region;
 - 20 f) a fifth adenovirus region from about base pair 30818 to about base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, wherein said fifth region is joined to said fourth region if said fourth region is present, or said fifth is joined to said third region if said fourth region is not present; and
 - 25 g) a sixth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region;
- provided that at least one of said second, third, and fifth regions is from Ad6.

30

59. The recombinant nucleic acid of claim 57, wherein said vector consists of:
- a) a first adenovirus region from about base pair 1 to about base pair 450 corresponding to either Ad5 or Ad6;

- b) a second adenovirus region from about base pair 3511 to about base pair 5548 corresponding to Ad5 or from about base pair 3508 to about base pair 5541 corresponding to Ad6, joined to said first region;
- c) a third adenovirus region from about base pair 5549 to about
5 base pair 28133 corresponding to Ad5 or from about base pair 5542 to about base pair 28156 corresponding to Ad6, joined to said second region;
- d) said gene expression cassette in a E3 parallel or E3 anti-parallel orientation joined to said third region;
- e) a fourth adenovirus region from about base pair 30818 to about
10 base pair 33966 corresponding to Ad5 or from about base pair 30789 to about base pair 33784 corresponding to Ad6, joined to said gene expression cassette; and
- f) a fifth adenovirus region from about base pair 33967 to about base pair 35935 corresponding to Ad5 or from about base pair 33785 to about base pair 35759 corresponding to Ad6, joined to said fourth region;
- 15 provided that at least one of said second, third, and fourth regions is from Ad6.

1/92

```

1      MAPITAYSQQ TRGLLGCIIT SLTGRDKNQV EGEVQVQVSTA TQSFLATCVN
51     GVCWTVYHGA GSKTLGAPKG PITQMYTNVD QDLVGWQAPP GARSITPCTC
101    GSSDLYLVTR HADVIPVRRR GDSRGSLLSP RPSVSYLKSS GGPLCPSPGH
151    AVGLFRAAUC TRGVAKAVDF VVESMETTM RSPVFTDNSS PPAVPQSFQV
201    AHLHAPTGS GSKTKVPAAYA AQGYKVLVLN PSVAATLGFG AYMSKAHGID
251    PNIRGTGVRTI TTGAPVITYST YGKFLADGGC SGGAYDIIIC DECHSTDSTT
301    ILGIGTVLDQ AETAGARLVV LATATPPGSV TVPHNIEEV ALSNTEIIFV
351    YGKAIPIEAI RGGRHILFCH SKKKCELA LKSLGLINAV AYYRGLDVSV
401    IPTIGDVVVV ATDALMTGYT GDFDSVIDCN TCVTQTVDFFS LDPTFTIETT
451    TVPQDAVSRS QRRGRTRGR RGIYRFVTPG ERPSGMFDSS VLCBCYDAGC
501    AWYELTPAET SVRLRAYLNT PGLPVCQDHL EFWESVFTGL THIDAHFLSQ
551    TKQAGDNFFY LVAYQATVCA RAQAPPSPSD QMWKCLIRLK PTLHGPTPLL
601    YRLGAVQNEV TLTHPTIKYI MACMSADLEV VTSTWVLVGG VLAALAAACL
651    TTGSVWVIGR IILSGRPAIV PDREFLYQEF DEMEEECASHL PYIEQGMQLA
701    EQFKQKALGL LQTATKQAEA AAPVVESKWR ALETFWAKHM WNFISGIQYL
751    AGLSTLPGNP AIASLMAPTA SITSPLTQOS TLLFNILGGW VAAQLAPPSA
801    ASAFVAGAGIA GAAVGSIGLG KVLVDILAGY GAGVAGALVA FKVMSGEMPS
851    TEDLVNLLPA ILSPGALVVG VVCAAILRRH VGPGBGAVQW MNRLIAFASR
901    GNVHVSPTHYV PESDAAARVT QILSSLTITQ LLKRLHQWIN EDCSTPCSGS
951    WLRDVWDWIC TVLTDFTKWL QSKLLPQLPG VPPFSCQRGY KGVWRGDGIM
1001   QTTCPGCAQI TGHVKNGSMR IVGPKTCSNT WHGTFFPINAY TTGPCTPSPA
1051   PNYSRALWRV AAEEYVEVTR VGDFHYVTGM TTDNVKCPQC VPAPPEFFTEV
1101   DGVRHLRYAP ACRPLLREEV TFQVGLNQYL VGSQLPCEPE PDVAVLTSML
1151   TDPSHITAET AKRRLARGSP PSLASSASQ LSAPSLKATC THHVSPDAD
1201   LLEANLLWRQ EMGGNITRVE SENKVVVLDS FDPLRAEED REVSVPAEIL
1251   RSKSKFFPAM PIWARPDYNP PLESWKDPD YVPFVVHGCP LPPIKAPPPI
1301   PPRKRRTVVL TESSVSSALA ELATKTFGSS ESSAVDSGTA TALPDQASDD
1351   GDKGSDESVEY SSMPPLEGE PGPDLSDGSW STVSEASED VCCSMSYTW
1401   TGALITPCAA EESKLPINAL SNSLLRHHNM VYATTSRSAG LRQKKVTFDR
1451   LQVLDDHYRD VLKEMKAKAS TVKALLSVE EACKLTTPHS AKSKFGYGAK
1501   DVRLNLSKAV NHIHSVWKDL LEDTVTPIDT TIMAKNEVFC VQPEKGRKP
1551   ARLIVFPDLG VRVCEKMALY DVVSTLPQVV MGSSYGFQYS PQQRVEFLVN
1601   TWKSKKNPMG FSYDTRCFDS TVTENDIRVE ESIYQCDDLA PEARQAISKSL
1651   TERLYIGGPI TNSKGQNCGY RRCRASGVLT TSCGNLTLCY LKASAACRAA

```

FIG. 1A

2/92

1701	KLQDCTMLVN	AAGLVVICES	AGTQEDAASL	RVFTEAMTRY	SAPPGDPPQF
1751	EYDLELITSC	SSNVSAHDA	SGKRVYYLTR	DPTTPLARAA	WETARHTPVN
1801	SNLGNIIIMYA	PTLWARMILM	THFFSILLAQ	EQLEKALDCQ	IYGACYSIEP
1851	LDLPQIIERL	HGLSAFSLHS	YSPGEINRVA	SCLRKLGVPF	LRVVRHRARS
1901	VRARLLSQGG	RAATCGKYL	NWAVKTKLKL	TPIPAASQLD	LSGNFVAGYS
1951	GGDIYHSLSR	ARPRWFMLCL	LLLSVGVGII	LLFNR	

FIG. 1B

3/92

```

1      GCCACCATGG CGCCCATCAC GGCCTACTCC CAACAGACGC GGGGCTACT
51     TGGTTGCATC ATCACTAGCC TTACAGGCCG GGACAAGAAC CAGGTCGAGG
101    GAGAGGTTCA GGTGGTTTCC ACCGCAACAC AATCCTTCCT GGCACACTGC
151    GTC AACGGCG TGTGTTGGAC CGTTTACCAT GGTGCTGGCT CAAAGACCTT
201    AGCCGGCCCA AAGGGGCCAA TCACCCAGAT GTACACTAAT GTGGACCAGG
251    ACCTCGTCGG CTGSCAGGCG CCCCCCGGGG CGCGTTCCTT GACACCATGC
301    ACCTGTGGCA GCTCAGACCT TTA CTGGTC ACGAGACATG CTGACGTCAT
351    TCCGCTGCGC CGGCGGGGCG ACAGTAGGGG GAGCCTGCTC TCCCCAGGC
401    CTCTCTCCTA CTTGAAGGCG TCTTCGGGTG GTCCACTGCT CTGCCCTTCG
451    GGGCACGCTG TGGGCATCTT CCGGGCTGCC GTATGCACCC GGGGGGTTGC
501    GAAGGCGGTG GACTTTGTGC CCGTAGAGTC CATGGAACT ACTATGCGGT
551    CTCGCTCTT CACGGACAAC TCATCCCCC CGGCGGTACC CGAGTCATTT
601    CAAGTGCGCC ACCTACAGC TCCCACTGGC AGCGCAAGA GTACTAAAGT
651    GCCGSCGTCA TATGCAGCCC AAGGGTACAA GGTGCTCGTC CTCAACTCGT
701    CCGTGTGCCG TACCTTAGGG TTTGGGCGT ATATGTCTAA GGCACACGCT
751    ATTGACCCCA ACATCAGAAC TGGGGTAAGG ACCATTACCA CAGGCGCCCC
801    CGTCACATAC TCTACCTATG GCAAGTTCTT TGCGATGGT GGTGTGCTCTG
851    GGGGCGCTTA TGACATCATA ATATGTGATG AGTGCCATTC AACTGACTCG
901    ACTACAATCT TGGGCATCGG CACAGTCTTG GACCAAGCGG AGACGGCTGG
951    AGCGCGGCTT GTCGTGCTCG CCACCGCTAC GCCTCCGGA GTGGTCACCG
1001   TGCCACACCC AACATCGAG GAGGTGGCCC TGCTAATAC TGGAGAGATC
1051   CCTTCTATG GCAAAGCCAT CCCATTGAA GCCATCAGGG GGGGAAGGCA
1101   TCTCATTTTC TGTCATTCCA AGAAGAAGTG CGACGAGCTC GCCGCAAAGC
1151   TGTGAGSCTT CGGAATCAAC GCTGTGGCGT ATTACCGGG GCTCGATGTC
1201   TCCGTCATAC CAACTATCGG AGACGTCGTT GTCGTGGCAA CAGACGCTCT
1251   GATGACGGGC TATACGGGCG ACTTTGACTC AGTGATCGAC TGTAAACATG
1301   GTGTCACCCA GACAGTCGAC TTCAGCTTGG ATCCACCTT CACCATTTAG
1351   ACGACGACCG TGCCFCAAGA CGCAGTGTG CGCTCGCAGC GGGCGGGTAG
1401   GACTGGCAGG GGTAGGAGAG GCATCTACAG GTTTGTGACT CCGGAGAAAC
1451   GGCCCTCGGG CATGTTGAT TCCTCGGTCC TGTGTGAGTG CTATGACGCG
1501   GGCTGTGCTT GGTACGAGCT CACCCCGGCC GAGACCTCG TTAGGTTGCG
1551   GGCTACCTG AACACACGAG GGTGCCCCGT TTGCCAGGAC CACCTGGAGT
1601   TCTGGGAGAG TGTCTTACA GGCTCACCC ACATAGATGC ACATCTCTTG
1651   TCCCAGACCA AGCAGGCAGG AGACAACCTC CCTTACCTGG TAGCATACCA

```

FIG. 2A

4/92

1701 AGCCACGGTG TGCGCCAGGG CTCAGGCCCC ACCTCCATCA TGGGATCAAA
 1751 TGTGGAAGTG TCTCATACGG CTGAAACCTA CGCTGCACGG GCCAACACCC
 1801 TTGCTGTACA GGTGGGAGC CGTCCAAAT GAGGTCAACC TCACCCACCC
 1851 CATAACCAAA TACATCATGG CATGCATGTC GGCTGACCTG GAGGTGCTCA
 1901 CTAGCACCTG GGTGCTGGTG GCGGAGTCC TTGCAGCTCT GGGCGGATAT
 1951 TGCCTGACAA CAGGCAGTGT GGTCAATTGT GGTAGGATTA TCTTGTCCGG
 2001 GAGGCCGGCT ATTGTTCCCG ACAGGGAGTT TCTCTACCAG GAGTTCGATG
 2051 AAATGGAAGA GTGCGCCTCG CACCTCCCTT ACATCGAGCA GGAATGCAG
 2101 CTCGCGAGC AATTCAAGCA GAAAGCGCTC GGGTTACTGC AAACAGCCAC
 2151 CAAACAAGCG GAGGCTGCTG CTCCTGTGGT GGAGTCCAAG TGGCGAGCCC
 2201 TTGAGACATT CTGGGCGAAG CACATGTGGA ATTTCATCAG CGGGATACAG
 2251 TACTTAGCAG GCTTATCCAC TCTGCCTGGG AACCCCGCAA TAGCATCATT
 2301 GATGGCATTC ACAGCCTCTA TCACCAGCCC GCTCACACC CAAGTACCC
 2351 TCTGTTTAA CATCTTGGGG GGGTGGGTGG CTGCCAACT CGCCCCCCCC
 2401 AGCGCCGCTT CGGCTTTGCT GGGCGCCGGC ATCGCCGGTG CGGCTGTGCG
 2451 CAGCATAGGC CTTGGGAAGG TGCTTGTGGA CATTCGTGCG GGTATGGAG
 2501 CAGGAGTGGC CGGCGCGCTC GTGGCCTTCA AGGTCAATGAG CGGCGAGATG
 2551 CCCCTCCACG AGGACCTGCT CAATCTACTT CCTGCCATCC TCTCTCCTGG
 2601 CGCCTGGTC GTCGGGGTGG TGTGTGCAGC AATACTGGGT CGACACGTGG
 2651 GTCGGGAGA GGGGGCTGTG CAGTGGATGA ACCGGCTGAT AGCGTTCGCC
 2701 TCGCGGGGTA ATCATGTTTC CCCACGCAC TATGTGCTTG AGAGCGACGC
 2751 CGCAGCGCGT GTTACTCAGA TCCTCTCCAG CCTTACCATC ACTCAGCTGC
 2801 TGAAAAGGCT CCACCAGTGG ATTAATGAAG ACTGCTCCAC ACGGTGTTCC
 2851 GGCTCGTGGC TAAGGGATGT TTGGGACTGG ATATGCACGG TGTTGACTGA
 2901 CTTCAAGACC TGGCTCCAGT CCAAGCTCCT GCCGCAGCTA CGGGAGTCC
 2951 CTTTTCCTC GTGCCAAGC GGGTACAAGG GAGTCTGGCG GGGAGACGGC
 3001 ATCATGCAAA CCACCTGCCC ATGTGGAGCA CAGATCACCG GACATGTCAA
 3051 AAACGGTTCC ATGAGGATCG TCGGGCTTAA GACCTGCAGC AACACGTGGC
 3101 ATGGAAACAT CCCCATCAAC GCATACACCA CGGGCCCTTG CACACCCTCT
 3151 CCAGCGCCAA ACTATTCTAG GCGCTGTGG CGGGTGGCCG CTGAGGAGTA
 3201 CGTGGAGGTC ACGGGGTGG GGGATTCCA CTACGTGACG GGCATGACCA
 3251 CTGACAACGT AAAGTGCCCA TGCCAGGTTC CGGCTCCTGA ATTCTTCAGC
 3301 GAGGTGGAGC GAGTGCGGTT GCACAGGTAT GCTCCGGCGT GCAGCCCTCT
 3351 CCTACGGGAG GAGGTACAT TCCAGGTGCG GCTCAACCAA TACCTGTTG

FIG. 2B

5/92

```

3401  GGTACACAGT ACCATGCGAG CCCGAACCGG ATGTAGCAGT GCTCACTTCC
3451  ATGCTCACCG ACCCCTCCCA CATCACAGCA GAAACGGCTA AGCGTAGGTT
3501  GGCCAGGGGG TCTCCCCCTT CCTTGGCCAG CTCCTCAGCT AGCCAGTGTG
3551  CTGCGCCTTC CTTGAAGGCG ACATGCACTA CCCACCATGT CTCTCCGGAC
3601  GCTGACCTCA TCGAGGCCAA CCTCCTGTGG CGGCAGGAGA TGGGCGGGAA
3651  CATCACCCGC GTGGAGTCGG AGAACAAGGT GGTAGTCCTG GACTCTTTTCG
3701  ACCCGCTTCG AGCGGAGGAG GATGAGAGGG AAGTATCCGT TCCGCGGGAG
3751  ATCTTGCGGA AATCCAAGAA GTTCCCCGCA GCGATGCCCA TCTGGGCGCG
3801  CCCGGATTAC AACCCCTCAC TGTTAGAGTC CTGGAAGGAC CCGGACTACG
3851  TCCCTCCGGT GGTGCACGGG TGCCCCGTGC CACCTATCAA GGCCCCTCCA
3901  ATACCACCTC CACGGAGAAA GAGGACGGTT GTCTTAACAG AGTCTCTCGT
3951  GTCTTCTGCC TTAGCGGAGC TCGTACTATA GACCTTCGGC AGCTCCGAAT
4001  CATCGGCCGT CGACAGCGGC ACGGCGACCG CCCTTCCTGA CCAGGCCCTCC
4051  GACGACGGTG ACAAGGATC CGACGTTGAG TCGTACTCCT CCATGCCCCC
4101  CTTTGGGGGG GAACCGGGG ACCCCGATCT CAGTGACGGG TCTTGGTCTA
4151  CCGTGAGCGA GGAAGCTAGT GAGGATGTCG TCTGCTGCTC AATGTCTTAC
4201  ACATGGACAG GCGCCTTGAT CACGCCATCG GCTGCGGAGG AAAGCAAGCT
4251  GCCCCATCAAC GCGTTGAGCA ACTCTTTGCT GCGCCACCAT AACATGGTTT
4301  ATGCCACAAC ATCTCGCAGC GCAGGCCTGC GGCAGAAGAA GGTCACCTTT
4351  GACAGACTGC AAGTCTTGA CGACCACTAC CGGGACGTGC TCAAGGAGAT
4401  GAAGGCGAAG GCGTCCACAG TTAAGGCTAA ACTCCTATCC GTAGAGGAAG
4451  CCTGCAAGCT GACGCCCCCA CATTCGGCCA AATCCAAGTT TGGCTATGGG
4501  GCAAAGGACG TCCGGAACTT ATCCAGCAAG GCGTTAACC ACATCCACTC
4551  CGTGTGGAAG GACTTGCTGG AAGACACTGT GACACCAATT GACACCACCA
4601  TCA/TGGCAAA AAATGAGGTT TCTGTGTCC AACAGAGAA AGGAGGCCGT
4651  AAGCCAGCCC GCCTTATCGT ATTCCAGAT CTGGGAGTCC GTGTATGCGA
4701  GAAGATGGCC CTCTATGATG TGGTCTCCAC CCTCCTCAG GTGCTGATGG
4751  GCTCCTCATA CGGATTCCAG TACTCTCCTG GGCAGCGAGT CGAGTTCCCTG
4801  GTGAATACCT GGAATCAAA GAAAAACCC ATGGGCTTTT CATATGACAC
4851  TCGCTGTTTC GACTCAACGG TCACCGAGAA CGACATCCGT GTTGAGGAGT
4901  CAATTTACCA ATGTTGTGAC TTGGCCCCCG AAGCCAGACA GGCCATAAAA
4951  TCGCTCACAG AGCGGCTTTA TATCGGGGGT CCTCTGACTA ATTCAAAGG
5001  GCAGAACTGC GGTATTCGCC GGTCCCGCSC GAGCGGCGTG CTGACGACTA
5051  GCTGCGGTAA CACCTCACA TGTACTTGA AGGCCTCTGC AGGCTGTGCA

```

FIG. 2C

6/92

5101 GCTGCGAAGC TCCAGGACTG CACGATGCTC GTGAACGCCG CCGGCCCTTGT
5151 CGTTATCTGT GAAAGCGCGG GAACCAAGA GGACGCGCGG AGCCTACGAG
5201 TCTTCACGGA GGCTATGACT AGGTACTCTG CCCCCCCGG GGACCCGCC
5251 CAACCAAGAAT ACGACTTGA GCTGATAACA TCATGTTCTT CCAATGTGTC
5301 GGTGCGCCAC **GATGCATCAG** GCAAAAGGGT GTACTACCTC ACCCGTGATC
5351 CCACCAACCC CCTGCGACGG GCTGCGTGGG AAACAGCTAG ACACACTCCA
5401 GTTAACCTCT GGCTAGGCAA CATTATCATG TATGCGCCCA CTTTGTGGGC
5451 AAGGATGATT CTGATGACTC ACTTCTTCTC CATCCTTCTA GCACAGGAGC
5501 AACTTGAAAA AGCCCTGGAC TGCCAGATCT ACGGGGCCCTG TTACTCCATT
5551 GAGCCACTTG ACCTACCTCA GATCATTGAA CGACTCCATG GCCTTAGCGC
5601 ATTTTCACTC CATAGTACT CTCCAGGTGA GATCAATAGG GTGGCTTCAT
5651 GCCTCAGGAA ACTTGGGTA CCACCTTGC GAGTCTGAG ACATCGGGCC
5701 AGGAGCGTCC GCGTAGGCT ACTGTCCAG GGGGGGAGG CCGCCACTTG
5751 TGGCAAGTAC CTCTTCAACT GGGCAGTGAA GACCAAACTC AAACCTACTC
5801 CAATCCCGGC TCGTCCCAG CTGGACTTGT CCGGCTGGTT CGTTGCTGGT
5851 TACAGCGGGG GAGACATATA TCACAGCCTG TCTCGTGCCC GACCCCGCTG
5901 GTTCATGCTG TGCTACTCC TACTTTCTGT AGGGGTAGGC ATCTACCTGC
5951 TCCCCAACCG ATAAA

FIG. 2D

7/92

```

1      GCCACCATGG  CCCCCATCAC  CGCCTACAGC  CAGCAGACCC  GCGGCCTGCT
51     GGGCTGCATC  ATCACCAGCC  TGACCGGCCG  CGACAAGAAC  CAGGTGGAGG
101    GCGAGGTGCA  GGTGGTGAGC  ACCGCCACCC  AGAGCTTCCT  GGCCACCTGC
151    GTGAACGGCG  TGTGCTGGAC  CGTGTACCAC  GGCGCCGGCA  GCAGACCCCT
201    GGCCGGCCCC  AAGGGCCCCA  TCACCCAGAT  GTACACCAAC  GTGGACCAGG
251    ACCTGGTGGG  CTGGCAGGCC  CCCCCCGCG  CCGCAGCCT  GACCCCTGC
301    ACCTGCGGCA  GCAGCGACCT  GTACCTGGTG  ACCCGCCACG  CCGACGTGAT
351    CCCCCTGCGC  CGCCCGCGCG  ACAGCCGCG  CAGCCTGCTG  AGCCCCCGCC
401    CCGTGAGCTA  CCTGAAGGGC  AGCAGCGGCG  GCCCCTGCT  GTGCCCCAGC
451    GGCCACGCGC  TGGGCATCTT  CCGCGCCGCC  GTGTGCACCC  GCGCGCTGGC
501    CAAGGCCGTG  GACTTCGTGC  CCGTGGAGAG  CATGGAGACC  ACCATGCGCA
551    GCCCGCTGTT  CACCGACAAC  AGCAGCCCCC  CCGCGTGCC  CCAGAGCTTC
601    CAGGTGGCCC  ACCTGCACGC  CCCACCGGC  AGCGGCAAGA  GCACCAAGGT
651    GCCCGCGGCC  TACGCGCGCC  AGGGCTACAA  GGTGCTGGTG  CTGAACCCCA
701    GCGTGGCCGC  CACCCCTGGG  TTCGGCGCCT  ACATGAGCAA  GGCCACCGGC
751    ATCGACCCCA  ACATCCGCAC  CGGCGTGCGC  ACCATCACCA  CCGCGCCCCC
801    CGTGACCTAC  AGCACCTACG  GCAAGTTCCT  GGCCGACGGC  GGCTGCACGG
851    GCGCGCCTA  CGACATCATC  ATCTGCGACG  AGTGCCACAG  CACCGACAGC
901    ACCACCATCC  TGGGCATCGG  CACCGTGTG  GACCAGGCCG  AGACCGCCGG
951    CGCCCGCCTG  GTGTTGCTGG  CCACCGCCAC  CCCCCCGGCG  AGCGTAGACG
1001   TGCCCCACCC  CAACATCGAG  GAGGTGGCCC  TGAGCAACAC  CGCGGAGATC
1051   CCTTCTACG  GCAAGGCCAT  CCCATCGAG  GCCATCCGCG  GCGCGCCCA
1101   CTTGATCTTC  TGCCACAGCA  AGAAGAAGTG  CGACGAGCTG  GCGGCCAAGC
1151   TGAGCGGCCT  GGGCATCAAC  GCCGTGSCCT  ACTACCGCG  CTTGGACGTG
1201   AGCGTGATCC  CCACCATCGG  CGACGTGGTG  GTGTGGCCA  CCGACGCCCT
1251   GATGACCGGC  TACACCGCGC  ACTTCGACAG  CGTGATCGAC  TGCAACACCT
1301   GCGTGACCCA  GACCGTGGAC  TTCAGCTGG  ACCCCACCTT  CACCATCGAG
1351   ACCACCACCG  TGCCCCAGGA  CGCCGTGAGC  CGCAGCCAGC  GCGCGGCCG
1401   CACCGGCCGC  GGCGCGCGCG  GCATCTACCG  CTTGTTGACC  CCGCGCGAGC
1451   GCGCCAGCGG  CATGTTTCGAC  AGCAGCGTGC  TGTGCGAGTG  CTACGACGCC
1501   GGCTGCGCCT  GGTACGAGCT  GACCCCGGCC  GAGACCAGCG  TGCGCCTGCG
1551   CGCTTACCTG  AACACCCCGG  GCCTGCCCGT  GTGCCAGGAC  CACCTGGAGT
1601   TCTGGGAGAG  CGTGTTCACC  GGCTGACCC  ACATCGACGC  CCATTCCTTG
1651   AGCCAGACCA  AGCAGGCCGG  CGACAACCTC  CCCTACCTGG  TGGCCTACCA

```

FIG. 3A

8/92

1701 GGCCACCGTG TGCGCCCGCG CCCAGGCCCC CCCCCCAGC TGGGACCAGA
 1751 TGTGGAAGTG CCTGATCCGC CTGAAGCCCA CCTGCACGG CCCCACCCC
 1801 CTGCTGTACC GCGTGGGCGC CGTGCAAGAC GAGGTGACCC TGACCCACCC
 1851 CATCACCAAG TACATCATGG CCTGCATGAG CGCCGACCTG GAGGTGGTGA
 1901 CCAGCACCTG GGTGTCTGGT GCGGCGGTGC TGGCCGCCCT GGCGCGCTAC
 1951 TGCTGACCA CCGGCAGCGT GGTGATCGTG GGCCGCATCA TCCTGAGCGG
 2001 CCGCCCCGCC ATCGTGCCCG ACCGCGAGTT CCTGTACCAG GAGTTGAGC
 2051 AGATGGAGGA GTGCGCCAGC CACCTGCCCT ACATCGAGCA GGGCATGCAG
 2101 CTGCGCCGAGC AGTTCAAGCA GAAGGCCCTG GGCTGCTGC AGACCGCCAC
 2151 CAAGCAGGCC GAGGCCGCGC CCCCCTGGT GGAGAGCAAG TGGCGCGCCC
 2201 TGGAGACCTT CTGGGCCAAG CACATGTGGA ACTTCATCAG CGGCATCCAG
 2251 TACCTGGCCG GCCTGAGCAC CCTGCCCGC AACCCCGCCA TGGCCAGCCT
 2301 GATGGCCTTC ACCGCCAGCA TCACCAGCCC CCTGACCACC CAGAGCACCC
 2351 TGCTGTTCAA CATCCTGGGC GGCTGGGTGG CCGCCAGCT GGCCCCCCCC
 2401 AGCGCCGCCA GCGCCTTCGT GGGCGCCGCG ATCGCCGCGC CGCGGTGGG
 2451 CAGCATCGGC CTGGCAAGG TGCTGGTGA CATCCTGGCC GGCTACGGGG
 2501 CCGCGCTGCG CGGCGCCCTG GTGGCCTTCA AGGTGATGAG CGGCAGATG
 2551 CCCAGCACCG AGGACCTGGT GAACTGCTG CCGCCATCC TGAGCCCCGG
 2601 CGCCCTGGTG GTGGCGGTGG TGTGCGCGC CATCCTGCGC CGCCACGTGG
 2651 GCCCCGGCGA GGGCGCCGTG CAGTGGATGA ACCGCTGAT CGCCTTCGCC
 2701 AGCGCGCGCA ACCACGTGAG CCCCACCCAC TACGTGCCCG AGAGCGACGC
 2751 CGCGCCCGC GTGACCCAGA TCCTGAGCAG CCGTGAACATC ACCCAGCTGC
 2801 TGAAGCGCCT GCACCACTGG ATCAACGAGG ACTGCAGCAC CCCCCTGCAGC
 2851 GGCAGCTGCG TGCGCGACGT GTGGGACTGG ATCTGCACCG TGCTGACCGA
 2901 CTTCAAGACC TGGCTGCAGA GCAAGCTGCT GCCCCAGCTG CCGCGGTGC
 2951 CCTTCTTCAG CTGCCAGCGC GGCTACAAGG GCGTGTGGCG CGGCGACGCG
 3001 ATCATGCAGA CCACCTGCC CTGCGGCGCC CAGATCACCG GCCACGTGAA
 3051 GAACGGCAGC ATGCGCATCG TGGGCCCCAA GACCTGCAGC AACACCTGGC
 3101 ACGGCACCTT CCCCATCAAC GCCTACACCA CCGGCCCTG CACCCCCAGC
 3151 CCGCCCCCA ACTACAGCG CGCCCTGTGG CGCGTGGCCG CCGAGGAGTA
 3201 CGTGGAGGTG ACCCGCTGG GCGACTTCCA CTACGTGACC GGCATGACCA
 3251 CCGACAACGT GAAGTGCCCC TGCCAGGTGC CCGCCCCGA GTTCTTCACC
 3301 GAGGTGGACG GCGTGCGCCT GCACCGCTAC GCGCCCGCCT GCGCCCCCT
 3351 GCTGCGCGAG GAGGTGACCT TCCAGGTGGG CTTGAACCAG TACCTGGTGG

FIG. 3B

9/92

3401 GCAGCCAGCT GCCCTGCCGAG CCCGAGCCCG ACGTGGCCGT GCTGACCAGC
 3451 ATGCTGACCG ACCCCAGCCA CATCACCAGC GAGACCGCCA AGCGCCGCCCT
 3501 GGCCCGCGCG AGCCCCCCCA GCCTGGCCAG CAGCAGCGCC AGCCAGCTGA
 3551 GCGCCCCCAG CCTGAAGGCC ACCTGCACCA CCCACCACGT GAGCCCCGAC
 3601 GCCGACCTGA TCGAGGCCAA CCTGCTGTGG CGCCAGGAGA TGGGCGGCAA
 3651 CATCACCCGC GTGGAGAGCG AGAACAAGT GGTGGTGTCTG GACAGCTTCTG
 3701 ACCCCCTGCG CGCCGAGGAG GACGAGCGCG AGGTGAGCGT GCCCGCCGAG
 3751 ATCTTGGCGA AGAGCAAGAA GTTCCCGCC GCCATGCCCA TCTGGGCCCC
 3801 CCCCAGCTAC AACCCCCCCC TGCTGGAGAG CTGGAAGGAC CCGGACTACG
 3851 TGCCCCCGCT GGTGCACGGC TGCCCCCTGC CCCCATCAA GGGCCCCCCC
 3901 ATCCCCCCCC CCCGCCGCAA GCGCACCGTG GTGCTGACCG AGAGCAGCGT
 3951 GAGCAGCGCC CTGGCCGAGC TGGCCACCAA GACCTTCGGC AGCAGCGAGA
 4001 GCAGCGCCGT GGACAGCGGC ACCGCCACCG CCCTGCCCGA CCAGGCCAGC
 4051 GAGCAGCGCG ACAAGGGCAG CGACGTGGAG AGCTACAGCA GCATGCCCCC
 4101 CCTGGAAGGC GAGCCCGCGC ACCCCGACCT GAGCGACCGC AGCTGGAGCA
 4151 CCGTGAGCGA GGAGGCCAGC GAGGACGTGG TGTGCTGCAG CATGAGCTAC
 4201 ACCTGGACCG GCGCCCTGAT CACCCCTGC GCGCCGAGG AGAGCAAGCT
 4251 GCCCATCAAC GCCCTGAGCA ACAGCTGTGT GCGCCACCAC AACATGTTGT
 4301 ACGCCACCAC CAGCGCAGC GCGGCCTGC GCCAGAAGAA GGTGACCTTC
 4351 GACCGCCTGC AGGTGCTGGA CGACCACTAC CGCGACGTGC TGAAGGAGAT
 4401 GAAGGCCAAG GCCAGCACCG TGAAGGCCAA GCTGCTGAGC GTGGAGGAGG
 4451 CCTGCAAGCT GACCCCCCCC CACAGCGCCA AGAGCAAGTT CGGCTACGGC
 4501 GCCAAGGACG TCGCAACCT GAGCAGCAAG GCGTGAACC ACATCCACAG
 4551 CGTGTGGAAG GACCTGCTGG AGGACACCGT GACCCCCATC GACACCACCA
 4601 TCATGGCCAA GAACGAGGTG TTCTGCGTGC AGCCCGAGAA GGGCGGCCCG
 4651 AAGCCCGCCC GCCTGATCGT GTTCCCGAC CTGGGCGTGC GCGTGTGCGA
 4701 GAAGATGGCC CTGTACGACG TGGTGAGCAC CCTGCCCCAG GTGTGTATGG
 4751 GCAGCAGCTA CGGCTTCAG TACAGCCCCG GCCAGCGCGT GGAGTTCTCTG
 4801 GTGAACACCT GGAAGAGCAA GAAGAACCCC ATGGGCTTCA GCTACGACAC
 4851 CCGCTGCTTC GACAGCACCG TGACCGAGAA CGACATCCGC GTGGAGGAGA
 4901 GCATCTACCA GTGCTGCGAC CTGGCCCCCG AGGCCCGCCA GGCCATCAAG
 4951 AGCTTGACCG AGCGCCTGTA CATCGGCGGC CCCCTGACCA ACAGCAAGGG
 5001 CCAGAACTGC GGTACCGCC GCTGCCGCGC CAGCGCGCTG CTGACCACCA
 5051 GCTCGGCGAA CACCCTGACC TGCTACCTGA AGGCCAGCGC CGCTGCCCCG

FIG. 3C

10/92

5101 GCGGCCAAGC TGCAGGACTG CACCATGCTG GTGAACGCCG CCGGCCTGGT
 5151 GGTGATCTGC GAGAGCGCCG GCACCCAGGA GGACGCCGCC AGCTGCGCG
 5201 TGTTCAACGA GGCCATGACC CGCTACAGCG CCCCCCCCG CGACCCCCC
 5251 CAGCCCGAGT ACGACCTGGA GCTGATCACC AGCTGCAGCA GCAACGTGAG
 5301 CGTGGCCCCAC GACGCCAGCG GCAAGCGCGT GTACTACCTG ACCCGCGACC
 5351 CCACCACCCC CCTGGCCCGC GCCGCCTGGG AGACGCCCGC CCAACCCCC
 5401 GTGAACAGCT GGCTGGGCAA CATCATCATG TACGCCCCCA CCTGTGGGC
 5451 CCGCATGATC CTGATGACCC ACTTCTTCAG CATCCTGCTG GCCCAGGAGC
 5501 AGCTGGAGAA GGCCCTGGAC TGCCAGATCT ACGGCGCCTG CTACAGCATC
 5551 GAGCCCCTGG ACCTGCCCA GATCATCGAG CGCTGCACG GCTGAGCGC
 5601 CTTACGCTG CACAGCTACA GCCCGGCGA GATCAACGC GTGGCAGCT
 5651 GCCTGCGCAA GCTGGGCGTG CCCCCCTGC GCGTGTGGCG CCACCGCGCC
 5701 CGCAGCGTGC GCGCCCGCCT GCTGAGCCAG GCGGCGCGCG CCGCACCTG
 5751 CGGCAAGTAC CTGTTCAACT GGGCCGTGAA GACCAAGCTG AAGCTGACCC
 5801 CCATCCCCGC CGCCAGCCAG CTGGACCTGA GCGGCTGGTT CGTGGCCGCC
 5851 TACAGCGCG GCGACATCTA CCACAGCCTG AGCCGCGCCC GCCCCGCTG
 5901 GTTCATGCTG TGCTGCTGC TGCTGAGCGT GGGCGTGGGC ATCTACCTGC
 5951 TGCCCAACCG CTAAA

FIG. 3D

11/92

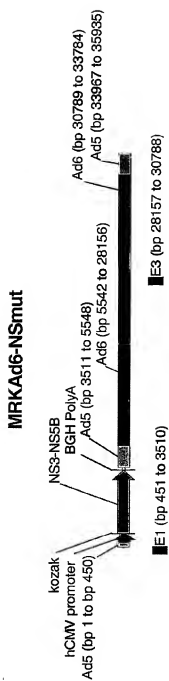


FIG. 4A

12/92

```

1  catcatcaat aatatacctt attttggatt gaagccaata tgataatgag ggggtggagt
61  ttgtgacgtg gcgcggggcg tgggaacggg gcggtgacg tagtagtggt gcggaagtgt
121 gatgttgcaa gtgtggcgga acacatgtaa gcgacgggat tggcaaaagt gacgtttttg
181 gtgtgcccgc gtgtacacag gaagtgcacaa ttttcgcgcg gttttaggcg gatgtttgag
241 taatttttgg cgtaaccgag taagattttg ccattttcgc ggggaaaactg aataagaggga
301 agtgaatctt gaataatttt gtgttactca tagcgcgtaa tattgtctca tggccgcgtg
361 gacttttgacc gtttacgttg agactcgccc aggtgttttt ctacagtggt ttccgcgttc
421 cgggtcaaaag ttgcggtttt attattatag gcgcccgcga tccattgcat acgttgtatc
481 catatcataa tatgtacatt tatatttgct catgtccaac attaccgccca tgttgacatt
541 gattattgac tagttattaa tagtaatcaa ttacggggtc attagtcat agcccatata
601 tggagtctcg cgttacataa cttacggtaa atggcccgc ttgctgaccg cccaacgacc
661 ccgcgccatt gacgtcaata atgacgtatg ttcccatagt aacgccaaata gggactttcc
721 attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt
781 atcatatgcc aaglacgcc cctattgacg tcaatgacgg taaatggccc gcctggcatt
841 atcccgagta catgacctta tgggactttc ctacttgcca gtacatctac gtattagtca
901 tctctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcgggttg
961 actcaacggg atttccaagt cttcccccac ttgacgtcaa tgggagtttg ttttggccc
1021 aaaaacaacg ggaactttca aaatgtcgta ccaactccgc cccattggcg caaattggcg
1081 gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg
1141 cttggagagc ccatccagcg tgttttgacc tccatagaag acacccggag cgtccagcc
1201 tccgcggccg ggaacggtgc attggaacgc ggaattcccg tgcacaagat gagatctgcc
1261 accatggcgc ccatcaagcg ctaactccaa cagacgcggg gctactctgg ttgcatctc
1321 actagcctta caggccggga caagaaccag gtgcaggggg aggttcagggt ttcttccacc
1381 gcaaacacaa ccttcctggc gacctgcgtc aagcgcgtgt cttgacacgt gtaccatgtt
1441 gctggctcaa agaccttagc cggcccaaag gcgcacaatca cccagatgta cactaatgtg
1501 gaccaggacc tctgcggctg gcaggcgccc ccgcggggcg cgttcattcc ggtgcgcggg
1561 tgtggcagct cagaccttta cttggtcaag cccagggcgt tctctcactt gaagggtctc
1621 cggggcgaca gtagggggag cctgctctcc cccagggcgt ccatcttcag ggaacttact
1681 tccggtggtc cactgctctg cctctcgggg cagcgtgtgg gcactctcag ggtccgcta
1741 tgcaaccggg gggttgcgaa ggcgttgagc tttgtgccc tagagtcacat ggaacttact
1801 atgcggtctc cgttcttcac ggacaactca aatccgtccg ttcccccggc cgtaccgca
1861 gtgcccacc tacacgtctc cactggcagc ggcaagagta ctaaatgccc ggcctgcata
1921 gcagcccaag ggtacaaggt gctcgtctcc aatccgtccg ttccgcctac cttagggttt
1981 ggggcgtata tgtctaaagg acacggtatt gccccaaca tcagaactgg gtaaggacc
2041 attacacacg gcgccccgt cactactctc acctatggca agtttcttgc gnatggtgtg
2101 tgctctgggg gcgcttatga catcataata tgtgatgagt gccattccaac tgactgcact
2161 acaatcttgg gcatcgccac agtctctggc caacgcggaga cggctggagc gcgcgttgtc
2221 gtgctcgcca cgtctacgcc tccgggatcg gtacacgtgc cacacccaaa catcgaggag
2281 gtggccctgt ctaatactgg agagatcccc tcttatggca aagccatccc catggaagcc
2341 atcagggggg gaaggcatct cattttctgt catctcaaga agaagtgcga cgactcgcc
2401 gcaaaagctg caggcctcgg aatcaacgct gtggcgattt acccggggct cagatgtgtc
2461 gtacacacaa ctatcggaga cgtcgttgtc gtggcaacag acgctctgat gcagggtcat
2521 acgggcgact ttgaactcagt gatcgactgt aacacatgtg ttaccacagc agtcgacttc
2581 agcttggatc ccaccttcac cattgagaag acgacggtgc ctcaagacgc agtgtcgccg
2641 tccagcgagg cgggtaggac tggcaggggt aggagaggca tctacaggtt tctgactccg
2701 gggagacggc cctcgggcat gttcgattcc tcggtctctg gtgagtgtcta tgaacgggg
2761 tgtgcttggg acgagctcac ccccgcggag acctcggtta ggttgcgggc ctactgcgac
2821 acacacaggt tgcccggttg ccaggaccac cttcttgttc gggagagtgt cttcaagac
2881 ctaccccaca tagatgcaca cttcttgttc agagccaagc aggcaggaga caactcccc
2941 taactcgttag cataccaagc cagcgttgtc gccagggtct aggcccccac tccactatgg
3001 gatcaaatgt ggaagtgtct catacggcgt aaactctacg tgacggggct gccacccctt
3061 ctgacacagg tgggagccgt ccaaatagag gtcaccctca cccaccctac aacccaatag
3121 atcatgggag ccatgtcggc tgacctggag gtcgtcacta gcactctggg cctgggtggc
3181 ggaagtcttg cagctctggc cgcgtattgc ctgacaacag gcagtggtgt catgtgggt
3241 aggattatct tgtccgggag gccggtattt gttccgcga gggagtctct ctaccaggag

```

FIG. 4B

13/92

```

3361 gccagcgaat tcaagcagaa agcgcctcggg ttaactgcaa cagccaccaa acaagcggag
3421 gcctagctgct ccgtggtgga gtccaaatgg cgagcccttg agacattctg ggcgaagcac
3481 atgttgaattc tcatacaggg gatacagtag ttagcaggct tatccaactct gctctggaaac
3541 cccgcgaatag catcattgat ggcattcaca gctctatoca ccagcccgct caccaccaaa
3601 agtaccctccc tggttaaacat ctgtggggggg tgggtggcgt cccaactcgc cccccccagc
3661 gcgcctctcgg ctttctgctgg ccgcggcgatc gccggtgcgg ctgttggcag catagggcctt
3721 ggggaaggctc ttgtggacat tctggcgggt tatggagcag gagtggccgg cgcgctcgtg
3781 gccttcaagg tcatacgagg cgagatgcgcc tccaccaggg acctgggtcaa actcggtcga
3841 gccatcctctc ctccctggcgc cctggttgctc ggggtcgtgt gtgcagcaat actcggtcga
3901 cagctgggttc cgggagagggg ggctgtgcagc tggatgaacc ggctgatagc gtctgcctcg
3961 ccggttgtaate atgtttccccc cagcactatc gtgcctgaga gcgacgcgcg agcgctgtgt
4021 actcagatccc tctccagcct taccatcact cagctgtcga aaaggctcca ccagtggatt
4081 aatgaagact gctccacacc gtgttccggc tegtgtctaa gggatgtttg ggactggata
4141 tgacagctgtg tgactgacct caagacctgg ctcacgtcca agctcctgcc gcagctaccg
4201 ggagtcctctt ttttctcgtg ccaacgcggg tacaaggagg tctggcgggg agacggcatc
4261 atgcacaacca cctgcccattg tggagcacag atcacccggc atgtcaaaaa cggttccatg
4321 aggatctctcg ggctaaagac ctgcagcaac acgtggcatc gaacattccc catcaacgca
4381 tacacacggg gccctcgcac acctcttcca gcgcacaaat attctagggc gctgtggcgt
4441 ttgtgcctcgt aggagtagct cgggttccagc cgggtggggg atttccaata cgttgacggc
4501 atgaccactg acaacgtaaa gtgcccatgc caggttccgg ctctcgaatt ctctacggag
4561 gtgcagcggag tgcggttgca caggtagctc ccgcggtgca ggccctctct ggctcctcct
4621 tctaacattcc aggtcgggct caaccaatac ctggttgggt cacagctacc atgcgagccc
4681 gaaccgggatg tagcagtgct cacttccatg ctacacgacc ctcccacatc cgcagcagaa
4741 acggtcaagc gtatggttgc cagggggtct ccccctcctc tggccagctc ttcagctagc
4801 cagtgtctcg cgcttctctc gaagcgacaa tgcactaccc acctatgctc tccggagcgt
4861 gacctatctg aggcacaacct cctgtggcgg caggagatgy gcgggaacat caccgcgtg
4921 gagtgcgaga acaagtygtt agtctcgacc tctttcgacc cgcttcgagc ggaaggagat
4981 gagaggaagc tatctgttcc ggcggagatc ctgcggaatc ccaagaagt ctcccgagc
5041 atgcccattct gggcgcgccc ggattacaa cctccactgt tagagtctc gtgaaggacc
5101 gactacgtccc ctcgggtgtt gcacgggtgc ccgttgcac ctatcaaggc cctccaata
5161 ccacctccac ggagaaagag gacggttgtc ctaacagagt cctccgttca tctccctta
5221 gcggagctcg ctactaaagc cttcggcagc tccgaatcat cggccgtcga cagcggcagc
5281 gcagccccc tctctgacca ggctctccgac gcggtgaca aaggatccga cgttgagtcg
5341 tactcttcca tgcccccctc tgagggggaa ccgggggacc ccgactcagc tgaaggtct
5401 tggctctaccg tgagcgagga agctagttag gatgtcgtct cgtctcaat gctccacga
5461 tggagagcgc ctttgatcac gccatgcgct gcggaggaaa gcaagctgcc catcaacgca
5521 ttgagcaact ctttgcctgc ccacataac atggtttatg ccacaacatc ctacagcgca
5581 ggctctgggc agaagaaggt cacttttgac agactgcaag tcttgagcga ccaactaccg
5641 gactgctcca aggagatgaa gcggaagcgc tccacagtta aggtataact cctactccta
5701 gaggaagcct gcaagctgac gcccccacat tccggcaaat ccaagtttgg ctatggggca
5761 aaggagcgtcc ggaacctatc cagcaaggcc gtaaacaca tccactccgt gtggaaggac
5821 ttgctgcgaag acactgtgac accaattgac accacatcac tggcaaaaaa tgaagtttct
5881 tgtgtccaac cagagaagag aggcgctaag ccagccgcct ttatcgtatt cccagactct
5941 ggaagtctgt tatgcgagaa gatggccctc tatgatgtgg tctccacctc tctcaagct
6001 gtgattggct cctcatacgg attccagtag tctctgggc agcgatcga tttcgtggg
6061 aataactcga aatacaagaa aaacccattg ggcttttcat atgacaactg ctgtttcgac
6121 tcaacggtca ccgagaacga catcctgtgt gagagatcaa ttaccaactg tttgtactc
6181 gcccccgaag ccagacagcg catataatcg ctacagagc ggctttatat cgggggtct
6241 ctgactaatt caaaaggcca gnaactcggt tatcgccggt gccgcgcgag cgcctgcgct
6301 acgactagct gcggttaaac cctcacatgt tacttgagg cctctgcagc ctctgcagct
6361 gcgaagctcc aggaactgcac gatgctcgtc aacgcgcgcg tccgttctgt tatctgtgaa
6421 agcgccggaa cccaagagga ccgcgcgagc ctacgagctc tccagggagc tatcactagg
6481 cactctgccc ccccggggg ccgcgcccaa ccagatatac acttggagct gataacatca
6541 tgttctccca atgtgtcgtt cgcacaagat gcataaggca aaagggtgta ctaactccc
6601 cgtgatccca ccaaccccc cgcacgggct cgttgggaaa cagctagaaa cactccagtt

```

FIG. 4C

14/92

```

6661 aactcctgcc taggcaacat tatcatgtat gcgcccaett tgtgggcaag gatgattctg
6721 atgactcaact tcctctccat cctcttagca caggagacaac ctgaaaaagc cctggactgc
6781 cagatctacag gggcctgtta ctccattgag ccacttgacc tacctcagat catgaaacga
6841 ctccatggccc tttagcgatt ttoactccat agttactctct caggttgaga caatagggtg
6901 gcttcatgcc tcaaggaaact tgggttacca cctctgcgag tctggagaca tggggccagg
6961 agctgcgcgc cttagcctaact gtcccagggg gggaggggccg ccaacttggtg caagtcaacc
7021 ttcaactgggg cagtgaagac caaactcaaa ctactccaaa tcccggctcg gtcccagctg
7081 gactttgcgc gctggttcgt tgctggttac agcgggggag acatatatca cagcctgtct
7141 cgtgcccgac cccgctgggt catgctgtgc ctactctac ttctgtagg ggtaggcctc
7201 tactctgtccc ccaaccggtta aatctagagc tgtgccttct agttgccagc catctgttgt
7261 ttgccctccc cccgtgccct ccttgaccct ggaaggtgcc actcccactg tccttctcta
7321 ataaaaatgag gaaattgcac cgcattgtct gagttagtgt catctatct tgggggtggg
7381 ggtgggcgag gacagcaagg gggaggattg ggaagacaat agcaggcatg ctgggggatgc
7441 ggtgggctct atggccgate ggcgcgcgtg actgaaatgt gtgggctggt cttaagggtg
7501 ggaagaataa tataagggtg ggttcttatg tagttttgta tctgttttgc agcagccgcg
7561 gccgccatga gcaccaactc ttgtgatgga agcattgtga gctcatattt gaacaagcgc
7621 atgcgcccat gggccgggggt cgtcagaat gtgatgggct ccagcatattg tgcctgcccc
7681 gtctcgccgc caaactctac tactttgacc taagagaccg tgtctggaaac cggctggctg
7741 actgcaagct ccgcgcgcgc ttccagccgtg gcagccaccg cccgcgggatg tgtgactgac
7801 ttgtgtttcc tgagcccgct tgcagacagt gcagcttccc gttccatccgc cgcgactgac
7861 aagttagcgg ctcttttggc acaattggat tccttgaccgc gggaaactaa tctcgtttct
7921 cagcagctgt tggatctgcg ccagcagggt tcgcctctga aggtcttccc cctcccaact
7981 gcggtttaa acataaataa aaaaccagac tctgtttgga ttgtgatcaa gcaagtgtct
8041 tgctgtctct atttaggggt ttggcgccgc cggtaggcgc gggaccagcg gtcctgtgcg
8101 ttgagggtccc tgggtatttt ttccaggagc tggtaaaagt gactctggat gttcagatg
8161 atgggcataa gcccgctctc tgggttgagt ggcaccact cagagacttc atgtcgcggg
8221 gtggttgtgt agatgatcca gtctgattgc caggggcagg cccttgtgtg aagtgtttac
8281 ttcagtatga agctgattgc caggggcagg agatgcattc tggactgtat ttttagttg
8341 agctgggatg ggtgcatacg tggggatatg ttcatgttgt gcagaaaccac
8401 gctatgttcc cagccataat cctccgggga agcttagaag gaaatgcgtg gaagaacttg
8461 tatcctgggc acttgggaaa ttgtctatgt atgcttagag ccaataatgat ggcgaatggg
8521 gagaagccct tgtgacctcc aagatthtcc atgcatctgt ctgggatcac taactgcata
8581 ccacggcgcg cggcctgggc agggcgtag taccctccac agatttgcat tccccacgt
8641 aggatgagat cgtcatagcc catttttaca aagcgcgggc ggaaggttgc agactcggtg
8701 ataatggttc catccggccc agggcgtag ttaccctcac agatttgcat tccccacgt
8761 ttgagtccag atggggggat catgtctaac tgcgggggga tgaagaaaaac ggtttccggg
8821 gtagggggaga tcaagtgga gaaagacagg tctctgagca gctgcgactc acgcgacgtc
8881 gtgggcccgt aatcaacac tataccggc tgaacttgtt ccttaagaga gctgcagctg
8941 ccgtcatccc tgagcagggg ggcacttcg ttaagcatgt cctgactctg agtttttccc
9001 ctgaccacaa ccgcagcaag cgcgtcgccg ccaagcgata gcagttctgt caaggagca
9061 aagtthttca acggtttag agctccgcgc gttagcatgc ttttagagct ttgccaagc
9121 agttccagcg ggtcccacag ctgcgtcaac tgcctcacgg tcatctgata cagcatatct
9181 cctcgtttcg cgggttgggg cggctttcgc tgcacggcag tagtctagc tagtccagc
9241 gggccagggt catgtcttcc caggcgcgca ggttcctcgt cagcgtagtc tgggttcagg
9301 tgaagggttg cgtccggggc tgcgcgctgc ccagggtgag ccttagagctg gctcgtgtg
9361 tgcgaagcg ctcocggctc tgcgcctgag tggcgccag cgttagcttg accatgggtg
9421 catagctcag cccctccgcg cgttggtccc tggcgccag cttgccttgc gaggagcgc
9481 ccgcaggagg gcatgcaga cttttgagg cgtagagctt gggcgcgaga aataccgatt
9541 ccggggagta ggcataccgc ccgcagcccc ccgagacggt ctgcattccc acgagccagg
9601 tgagctctgg ccgttcgggg tcaaaaacca ggtttccccc atgcttttgc atgcgtttct
9661 tactcttggt ttccatgagc cgggtgtccc gctcgttgag gaaaaggtct tccgtgtccc
9721 cgtatacaga cttgagagggc ctgtcctcga cgggtgttcc cgggtctccc tcgtatagaa
9781 actccgaccca cctcagacg aggcctcgcg tccaggccag caggaaggtg ctaaggtggg
9841 aggggtagcg gtcgttgtcc actagggggt ccactcgtc caggtgtgta agacacatg
9901 cgcctctctc ggcatacagg aaggtgattg gtttataggt tagagccagc tgaccgggtg

```

FIG. 4D

15/92

```

9961 ttcttgagg ggggtataa aaggggtgg gggcggttc gtctcaete tottcegat
10021 cgtgtgtgc gagggccagc agtactccet ctcaaaagcg gcatgactt
10081 ctgcgtcaag atgtgcagtt lccaaaaaac gatattcacc tggcccgagg
10141 tgatgccttt gaggytgccc gcgtccactt ggtcagaaaa gacaactctt ttgttggtaa
10201 ctgtgtgtgc aaacgacccg tagagggcgt tggacagaaa ctggcgatg tggcgacagg
10261 tttgggtttt gtgcgcatcg gcgcgtccct tggcgcgatg gtttagctgc acgtattcgc
10321 ggcgaacgca cgcgcattcg ggaagaacgg cgcgcgtcct tgggtcgctc gtccgggact aggtgcacgc
10381 gccaaacgcy gttgtgcagg gtgacaaagt caacgtcggt ggcctacctc ccgcgtaggc
10441 gctcgtttgt ccagcagagg cggccgcctc tgcgcgagca gaatggcggt agtgggtcta
10501 gctcgtcttc gtccgggggg tctgcgtcca cggtaaagac cccgggcgag cggcgcgctt
10561 cgaagtgttc tatcttgcat ccttgcaagt ctacgcctcg ctgccaatgc aggcgcgcaa
10621 ggcgcgcctc gtatggggtg agtgggggac ccataggcat ggggtggggt agcgcggagg
10681 cgtacatgcc gcaaatgtcg taacagtaga ggggtctctc gagtattcca agatatgtag
10741 ggtagcatct tccaccgcgg atgctggcgc gcacgtaatc gtagattcgc tgcgaggagg
10801 cgaggaggtc gggaccgagg ttgctacggc cgggctgctc tgctcggaag actatctgcc
10861 tgaagtgtgc atgtgagttg gatgatagg ttggacgctg gaagacgttg aagctggcgt
10921 ctgtgagacc tacgcgtcca cgcacgaagg agcgctagga gtccgcgagc gtcgcgcgag ttgttgacca
10981 gctcggcggt gacctgcacg tctaggggcg agtagtccag ggtttccctg tatgtgtcat
11041 acttactctc tccctttttt tctccacagt cgcggttgag gacaaactct tccgggttct
11101 tcaagtactc ttggatcgga aaccgctcgg cctccgaacg gtaagagcct agcatctaga
11161 actggttgac ggcctggtag gcgcagcatc ccttttctac ggytagcggc tatgcctgag
11221 cggccttcgc gacgcgagtg tgggtgagcg caaaggtgtc cctaaccact actttgaggt
11281 actggtatct aaagtcaagt tctgtcgcat cgcctcgtc ccagagcaaa aagtcgcgtc
11341 gcttttgga acgcgggttt ggcaggcgca aggtgacatc gttgaagagt atctttccgc
11401 cgcgagcatc aaagtlcgct gtgatcgga cacttcggcg cacttcggaa cggttgttaa
11461 ttacttggcg ggcgagcagc atctcgtaa agcctltgat gttgtggccc acaatgtaaa
11521 gttccaagaa gcgcgggatg ccttggatgg aaggcaattt ttaagltcc tcgtaggtag
11581 gctcttcagg ggaagctgagc ccgtgctctg aaagggccca gtcctgaaga tgagggttgg
11641 aagcgacgaa tgagctccac aggtcacggg ccattagcat ttgcaggtgg tcgcgaaagg
11701 tcttaaaactg gcgacatgat gccatttttt ctggggtgat gcagtagaag gtaagcgggt
11761 cttgttccca cgggtcccat ccaaggtccg cggctaggtc tcgcgcggcg gtccactagag
11821 gctcatctcc gccgaacttc atgacagca tgaaggcgac gagctgcttc ccaaaaggcc
11881 ccatccaagt ataggtctct acatcgtagg tgacaaagag acgctcgggt cyaggtgcg
11941 agcgcgatcg gaagaactcg atctcccgcc accagttgga ggaagtggctg ttgatgtggt
12001 gaaagttagaa gtccctcgca cggcgccaac actcgtgctg gcttttgtaa aaacgtgcgc
12061 agtactggca gcggtgcacg ggctgtacat cctgcacgag gttgacctga cgcaccgcga
12121 caaggaagca gagtgggaat ttgagccctc ccgcttcggc gtttggctgg ttgttcttca
12181 ctccggtcgc ttgtccttga ccgtctggct gctcgagggg agttacgggt gatcgacca
12241 ccacgcgcgc cgaagcccaa gtccagatgt cgcgcgcggc cggctcagg tcaggcgga
12301 catcgcgcag atgggagctg tccatggtct ggaagctccc cggcgtcagg tcaggcgga
12361 gctcctcgag gtttacctcg catagcggg tcaaggcgcg gctaggttcc aggtgatacc
12421 tgatttccag gggctggtg gtggcgcgct cgaatgggtg caaagagcgc catcccgga
12481 gcgcgaactc ggtacgcgc gcggcgcggt cggcgcgggg ggtgtcctgt gatgatcat
12541 ctaaaacgcy tgacgcggcg gggcccccgg aggtaggggg ggtcggggag ccgcgcggag
12601 agggcgagg ggcacgtcgg cgcgcgcgcg ggcgcgcgag tgggtctgag cgcggaggtt
12661 gctggcgacg cgcagcgcgc ggcggttgat ctctgaatc tggcgctctc gcgtgaagac
12721 gaocggcccg gtgagcttga acctgaaga gaggtcgaca gaataaattt aggtctgtt
12781 gaocggcgcc ttggcgcaaa tctcctgcac gtcctctgag ttgtttgat ggcgtatctc
12841 ggcctggaac tgcctgatct ctctcctctg gagatctccg cgtcccgctc gctccacctt
12901 gggcgcgagg tctgttgaga cttggcccat gagctgcgag aagcgtctga ggcctccgtc
12961 gttccagacg cggctgtaga ccacgcccc ttccgcatcg cggcgcgaga tgaccacctg
13021 cgcgcagattg agctccacgt gcggggcgaa gcggcgtagg ttccgacgc gcctgaagag
13081 gtatgttagg gtgggtggcg tgtgttctgc cacaagaagc tacataaacc agcgcgcac
13141 cgcgattcg ttgatatccc ccaaggctcc atggcctcgt aggaactca
13201 ggcgaagtgt aaaaactggg agtgcgcgc gcacacggtt aactcctcct ccagaagacg

```

FIG. 4E

16/92

13261 gatgagctcg ggcacagtggt cgcgcacctc gcgctcaaa gctacagggg cctcttcttc
 13321 ttcttcaatc ttctcttcca taaggggctc cctcttctct ttcttggcg gcgggggggg
 13381 aggggggaca cggcgggcac gacggcgcac cgggaggcgg tcgacaaagg gctcgatcat
 13441 ctcccccggc cgacggcgca tggcttcggg gacggcgcg cgttctcgc gggggcgacg
 13501 ttgaaagacg cgcgcgctca tgtcccggtt atgggttggc ggggggtcgc cgtcgggcag
 13561 ggggtttctc ctaacgatgc atctcaacaa ttgttctgta ggtactccgc caacggggga
 13621 cctgagcgag tcgcgcatga cgggatcggg aaacctctcg agaaaggcgt ctaaacagtc
 13681 acagtcgcaa ggttaggtga gcacgtggcg gggcggcagc gggcgcggt cggggttgtt
 13741 tctggcgagg gtgctgctga tgatgtaatt aaagtagggc gtccttgagac ggcggatggt
 13801 cgacagaagc accatgtcct tgggttcggc ctgctgaatg cgcaggcggt cggcctatgc
 13861 ccaggctctg ttttgacatc ggcgcaggtc tttgtagtag tcttgcabga gcccttctac
 13921 cgcactctct tcttctcctt cctcttgtcc tgcctctctt gcactatgt ctcggagcgc
 13981 ggcggagttt ggcgtaggtt ggcgcctctt tctctccatg cgtgtgaccc cgaagccctc
 14041 catcgctgga agcaggggca ggtcggcgac aacgcgctcg gctaatatgg cctcgtgcac
 14101 ctgcgtgagg gttagactga agtcgtccat gtccacaaag cgggtggtatg cgcctcgtt
 14161 gatgtgttaa tgcagttgg ccataacgga ccagttaacg gtctggtgac ccgctgcga
 14221 gauctcgggt tacctgagac gcgagtaagc ccttgagtca aagacgtagt cgttccaagt
 14281 ccgcacacgc tactggtatc ccaccaaaaa gtgcggcggc ggtcggcggt agaaaggcca
 14341 cgttaggggt gccggggctc cggggggcag gtcttccaac ataaggcgat gatattccga
 14401 gatgtacctg gatccacagg tgatccggcg ggcggtggtg gagggcgcg cgaagtcacg
 14461 gacgcgggtc cagatgttgc gcagcggcaa aaagtgtccc atggtcggga cgtctggccc
 14521 ggtcagcggc gcgcagtggt tgacgctcta gacgctgcaa aaggagagcc tcgaagcggc
 14581 cactcttccg tggctcgggt gataaattcg caagggtatc atgcggcagc accggggttc
 14641 gaacccccga tccggcctgc cgcgcctgac cgcgctgato catcggtta cgcgaacca
 14701 ggtgtgcgac gcttttttgg ccaactggcg cgcgcggcgt ttttccaaag gttgagtccg
 14761 tgcctgccta agtggtctgc cctgtctcgc cgggagcggc ggaactgcgc gaacccccc
 14821 gaaagcatta agtggtctgc cgggccccgc cggaaaacag gcagagcccc tttttgctt
 14881 gggacccccg gttcagatct ccaaatctct gcgccccctt cctctctctt acccggtcag ggcagggcaa
 14941 gtcatgcagc accccgcttg tgcggcagat gggcaccttc cctctctctt cgcggcgccc ggaacccgca
 15001 ttccagatg catccggtgc ggcacacccc cagatgttga ttacgaaccc ggagcgcccc
 15061 agagcagcgg cagacatgca gcgagggcct gcgcgggcta ggagcgcccc cctctgagcg
 15121 atccgcggct gacgcggcgg cgcagggcct gcgcgggcta taogtgcgcg ggcagaacct
 15181 ctacttggac ttggaggagg gcgagggcct ggcggggcta gacggaagt tccatgcagg
 15241 acacccaagg gtgcagctga agctgtgacac ggcggaggcg taogtgcgcg ggcagaacct
 15301 gtttcgcgac cgcgaggagg aggagcccga gggagatcgg gatcgaaagt tccatgcagg
 15361 cgcgagatgt cggcatggcc tgaacccgga gcggttctgt cgcggaggagg accttgagcc
 15421 cgaagcgcgc accgggatta gtcccgcgcg gcacacagtg gcgcccgcgc acctgtgaa
 15481 cgcgtacgag cagacggtga accaggagat taactttcaa aaaagcttta acacacaggt
 15541 gcgcagcctt tggcgcgcgc aggaagtggtg cgcgctcatg gcgcagctgt agcactttgt
 15601 aagcgcgcgt gagcaaaacc caaatagcaa gcgctcatg cgcagcagtg tctctatagt
 15661 gcagacacag agggacaacg aggcatttcag tgataaacat tctgcagagc atagtgtgtc
 15721 gggcgctgtg ctgctcgatt tggcgcgcac taactattcc atgtctagtc ttggcaagt
 15781 cttagcctg cgtgacaagg tggcgcgcac taactattcc atgtctagtc ttggcaagt
 15841 ttacgcccgc aagatatacc atacccttta gcttcccata gacaaggagg taagatctga
 15901 ggggtttctac atgcgcatgg cgttaaggtt cgttaacttg agcacagacc tggcggttta
 15961 tcgcaacgag cgcattccaca aggcgcgtgag cgtgagcggc cgcgcgcgag cgcgcgcgag
 16021 cgaagctgat cacagcctgc aaaggggcct cgtgagcagc ggcagcggcg atagagagc
 16081 cgaagctctac tttgacgcgg cgcgtgacct cgcgtggggc ccaagccgac cgcgccttga
 16141 ggcagctggg gcgcgacctg ggcctggcgtt ggcaaccgcg cgcgctggca acgtcgcgcg
 16201 cgtgaggagg tatgacgagg acgatgagta cggaccggag gacggtgagt actaagcgt
 16261 gatgtttctg atcatgatgt ccaagacgca cggaccggcg cgttcggcg cgcgctcttc
 16321 accagccgct ccggccttaa ctcacggcac gactggcgcc aggtcatgga cgcgcatcat
 16381 tgcgtctctg cgcgcaaccc tgacgcttgc cgcgacgagc cgcagggcaa cgcgctcttc
 16441 gcaattctgg aagcgggtgt cccggcgcg ccaaacccca cgcagagaa gctcgtggcg
 16501 atcgtaaacc cgttggcgga aaacaggggc atccggcccg atgaggccgg cctggtctac

FIG. 4F

17/92

```

16561 gacgcgctgc ttcagcgcgt ggcctgctaac aacacagcaga acgtgcagac caacctggac
16621 cggtctgtgct gggatgtgtgc ggcgcgcgtg agcgcgcgcga agcgcagggc
16681 aacctgtggct ccatggttgc actaaacgccc ttccctgagta cacagcccgcc caacgtgcgc
16741 cggggacacgg aggactacac caactttgtg agcgcactgc ggcctaattg gactgagaca
16801 ccgcaaatgt agtgtatca gtcccgggcca gactattttt aagaacttgc aggggctgtg gggggctgcg
16861 ctgcagacgg taaacctgag ccaggctttc gaccgtgtct agcttgcgta cgcaccaact gcgcctgtg
16921 gctcccacag gcgaccgcgc gaccgtgtct ggcagcagat ggcagcgctg atgtggagca atacctagg
16981 ctgctgtctaa cactgtacct cactgtacct cgaggccata ggtagggcgc ccggggacac atacctagg
17041 cacttgcctga cactgtacct cactgtacct cgaggccata ggtagggcgc ccggggacac atacctagg
17101 caggagatta caagtgttag ccgcgcgctg ggcgcagggag acacggggcg cctgggggca cctgggggca
17161 acctgaact acctgtctgac gcgctatgtg caaacgcgcg caaaaaatcc cctcgttgca cagttaaac
17221 agcgaggagg agcgattttt ggcgtatgtg atgacgcgcg cagcagagcg tgagccttaa cctgatgcg
17281 gacgggggtaa cgccacgcgt ggcgctggac atgacgcgcg gcaacatgga accgggcatg
17341 tatgcctcaa accggcgctt tatcaatcgc ctaattggact actgtcatcg cgcggcgccg
17401 gtgaaccccg agtatttcac caatgccatc ttgaaccgcg acttgctacc gccctcgtg
17461 ttctacacgg ggggatttga ggtgcccgag ggtaacgatg gatacctctg ggaacacata
17521 gacgacagcg tgttttccccc gcaacgcgcg accctgctag agttgcacaa acgcgcagcg
17581 gcagagcgcg cgctgcgaaa ggaagcttcc cgcagggccaa gcagcttgtc cgaatagggc
17641 cctgcggccc cgcggtcaga tgcctagtag ccatttccaa gcttgatagg gtctcttacc
17701 agcaactcgca ccaccgcgcc gcgctgtgct ggcgcagggag agtacctaaa caactcgtg
17761 ctgcagcgcg agcgcgaaaa gaaactgcct cccggctttc ccaacaacg gcgcagggac
17821 ctagtggaca agatgagtag atggaagacg tatgcgcagc agcacaggga tgtcccgccg
17881 ccgcgcgcgc ccaccgcgtc tcaaaaggca cagcgtcagc ggggctcgtg gtgggaggac
17941 gatgactcgc cagacgacag cagcgtcttg gatctgggag ggaagtgcac cccgtttgca
18001 cactctgcgc ccaggctggg gagaatgttt taataaaaaa catgatgcaa aataaaaaa
18061 tcaccaagcg catggcacgc agcgttgtgt ttcttgtatt ccccttagta tgcgcgcgcg
18121 ggcgatgtat gaggaaagtc ctctctctcc ctacgagagc gtggtgagcg cgcgcgcagt
18181 ggcgcgcgcg ctgggtttcac cctctgatgc tcccctggac ccgcgcgtcg tgcctccgcg
18241 gtacctgcgg cctaccgggg ggagaaacag cctccgttac tctgagttgg cacccttatt
18301 cgacacaccc cgtgtgtacc ttgtggacaa caagtcaacg gatgtggcat cctgaacta
18361 ccagaaacgc cacagcaact ttctaaccac ggtcattcaa aaactgact acagcccgcg
18421 ggaggcgaac acacagacca tcaatcttga cgcacggctc cactggggcg gcgacctgaa
18481 aaccatcctg cataccacaa tgccaaatgt gaacgagttc atgtttacca ataagtttaa
18541 ggcgcgggtg atggtgtcgc gctogcttac taaggacaaa caggtggagc tgaatacga
18601 gtgggtggag ttacgcgtgc ccgaggggcaa ctactccgag accatgacca tagaccttat
18661 gacaacacgc atcgtggagc actacttgaa agtgggcagc cagaacgggg ttctggaag
18721 cgacatcggg ttaaagtttg acaccgccaa ctacgactg gggtttgacc cagtcatcgc
18781 tcttgtcatg cctggggtat atacaacga agccttccat ccagacatca ttttgcgc
18841 aggatcgcg gtggacttca cccacgcgcg cctacgcaac ttgttggcca tccgcgaagc
18901 gcaacccctc caggaggcct ttaggtcac ctacgatgac ctggagggtg gtaaccttc
18961 cgcactgttg gatgtggaag cctaccagcg aagcttgaaa gatgacacc aacgagccgc
19021 ggttggcgca ggcggcgcca acaacagtgg cagcggcgcg gaagagaact ccaacgcctc
19081 agctgcgcga atgcagccgg tggaggacat gaacgatcat gccattcgcg gcgcacact
19141 ggcacacagg gcggaggaga agcgcgctga ggcgcaggca gcgcccgaag ctgcgcgcgc
19201 cgtctcgag cgtgcacac ccgaggctga gaangctcag aagaaacggg tgataaaac
19261 cctgacagag gacagcaaga aacgcagtta caacctaaat agcaatgaca gcaacctcac
19321 cagtaaccgc agctgttacc ttgcatacaa ctacggcgac cctcagcgcg ggtacacttc
19381 atggacctcg ctttgtacct ctgagctaac ctgcggctcg gagcagatat actggtcgtt
19441 gcccgacatg atgcaagacc ccgtgaacct cgcgtccacg cgcctccacg gcaacttcc
19501 ggttgtgggc gcgagctgt tgcccgctga cctcaagagc ttctacaag accagcgctg
19561 ctactcccg ctcactccgc agtttacctc tctgacccac gtttcaact gctttccga
19621 gaaccagatt ttggcgcgcc cgccagcccc caccatcacc acgctcagtg tctcagagt
19681 tgcctctaca gatcacggga cgtaccgcgt gcgcacagc atcgaggagc ttggcagagt
19741 gaccattact gacgcagag ccgcacctgt cccctacgtt tacaaggccc tggcattagt
19801 ctgcgcgcgc gtcctatcga gcgcacctt ttgagcaacg atgtccatcc ttatatcgcc

```

FIG. 4G

18/92

19861 cagcaataaac acaggctggg gcctgcgctt cccaagcaag atgttttggcg gggccaagaa
 19921 ggcactccgac caacaccag tgcgcgtgcy cggcgactac cgcgcgcctt gggcgcgca
 19981 caaacgcggc cgcactggcg gcaccacogt cgtatgacgc atcgacgcgg tggtaggaga
 20041 ggcgcgcaac tacacgccc cgcgcgcggc agtgctccacc gtggacgcgg ccatcagac
 20101 cgtgctgcgc ggaagccgcg gctacgctaa aatgaagaga cggcggaggc gctgtagca
 20161 tggcgtgcgc cgcgcaccgc gcactgcgcg ccaacgcgcg cggcgcggcg gctttaaccg
 20221 cgcacgtgcg accgcggcac gggcggccat gcgagccgct cgaaggctgg ccgcgggtat
 20281 tctcaactgt ccccccagg ccaaggacgc agcggccgcg gacagcgcg cggccattag
 20341 tgctatgact cagggtcgca ggggcaacgt gtaactgggt cgcgactcgg ttacggcgct
 20401 ggcgcgtgcc gtgcgcaccc gcccccgcg caactagatt gcaataaaaa actactaga
 20461 ctctgactgt tgatgtatc cagcggcggc ggcgcgcac gaagctatgt ccaagcgca
 20521 aatcaagaa gagatgctcc aggtcatcgc gccggagatc tatgcccccc cgaagaaggaa
 20581 agagcaggat tacaagcccc gaagctaaa cggggtcaaa aagaaaaaga aagatgatga
 20641 tgatgatgaa cttgacgac aggtggaaact gttgcacgcg accgcgccca ggcgcgggt
 20701 acagtggaaa ggtcgacgcg taagacgtgt tttgcgccc ggcaccaccg tagtctttac
 20761 gcccggtgag cgtctcaacc gcaactcaaa cgcgctgtat gatgaggtgt acggcgacga
 20821 ggaactgctt gacgaggcca acgagcgctt cggggagttt gctacgtaa agccgcataa
 20881 ggaactgctt gcgttgcgcg tggacgagg ccaaccaca cctagctcaa agcccgctac
 20941 actgcgacg gtgctgcggc cgcttgaccc gtcgaagaa aagcgcgcgc taagcgcga
 21001 gctgctgac ttggcaccga cgtgcagct gatggtacc aagcgtcagg tactggaaga
 21061 tgctttgaaa aaaatgacgc tggagcctgg gctggagccc gaggtccgcg tgcggccaat
 21121 caagcagctg gcaccgggac tggcgctgca gaccgtggac gttcaagata ccaccaccag
 21181 tagcaactgt attgcccgcg ccacagaggc catgagaca caaacgtccc cggttgctt
 21241 ggcgttgcca gatgcgcggc tgcacggcgc cgtcgcggcg gctgccaaga cctctacgga
 21301 ggtgcaaacg gacccgtgga tgtttcgtgt ttcaagcccc cggcgtccgc ccgcttcaag
 21361 gaagtacgca gccgcgcagg cgcactggc acactacgc ccccagaaga cgaagcaacta
 21421 tacccccgcg tatcgtggct acactacgc ccccgccag cgcgtctggt ccccgatttc
 21481 aacccaccat ggaacccgac gccgcgctgc ccgtgcgac cgcgtctggt cgaactggc
 21541 cgtgcgcagg gtggctcgcy aaggaggcag gacccctggg ctgccaacg cgcgtacaa
 21601 ccccagatc gtttaaaagc cggctcttgc ggttcttgca gatattggcc tcaactgcg
 21661 cctccgcttc ccggtgcggc gattccgagc aagaatgcac cgtaggaggc gctgcgcgg
 21721 ccaagcgcgt acggggcgga tgcgtcgtgc gcaccaaccg cggcggcgcg cgtcgacgc
 21781 tgcactgcgc ggcgggtatc ttgccctcct tatccactg atccgcgcg cgaatggcgc
 21841 cgtgcccgga attgcatccg tggccttgca ggcgcagaga cactgattaa aaacaggtta
 21901 catgtggaaa aatcaaaaata aaagcttgga ctctcacgct cgtttgtccc tgtaactatt
 21961 ttgtagaaat gaagacatac actttgcgtc actggcccg cgacacgggt cgcgcgcgtt
 22021 catgggaac ttgcaagata tgcgcacga caatagagc ggtggcgctt gtagtgagg
 22081 ctgcgtctgg agcggcataa aaaatttcgg ttccgcgctt aagaactatg cgaacaaagc
 22141 cgtgaacagc agcacaggcc agatcgtgag ttacaaagtg aagaagcaaa atttcaaca
 22201 aaagtggtga gatggcctgg cctctggcat taccggggtg gtggaactgg ccaacacaggc
 22261 agtgcaaaat aagattaaac gtaagcttga tagccgcctt cccgtagagg agctccacc
 22321 ggcgtgtgag acagtgtctc cagaggggag tggcgaaaag cgtccgcgac ccgacaggga
 22381 agaaaactgt gtgacgcaaa tagacgagcc tccctcgta c gaggagacac taagcgaag
 22441 cctgccacc acccgtccca tgcgcgccat ggcctaccga gtgctgggac agcacacca
 22501 cgttaacgct gacctgcctc ccccgcgca caccagcag aaacctgtgc gcccagccc
 22561 gtcgcgcgtt gttgtaaccc gtccactgac cgtgcgccc cgcgcgcgc ccagcggctc
 22621 gcgactcgtt cggcccgtag ccaatggca cctggcaaac gactcgtggg taagcgaag
 22681 ttgtgggggt caatccctga agcgcgcgac atgctctga tagctaaagt cgcctatgtg
 22741 tttcatgtat cggctcatgt cgcgcgcgca ggaactcgt agccgcgcgc gctcgtctg
 22801 ccaagatggc tacccttcg atgatgcgc agtggtctta catgcacatc tcggggcagg
 22861 agccctcgga gtacctgag cccggcgctg tgcagttcgc ccgcgcacc gagactact
 22921 tcagctgtaa taacaagttt agaaacccca cgttgcgcgc tacgcacgac gtgaccacag
 22981 acccgcttca cggtttgac ctcgggttca tcccgcgga ccgcgagata actgcttact
 23041 cgtacaaggc gcggttacc ctactgtggt gtgataaccg tgtctagac atgcttcca
 23101 cgtactttga catccgcggc gtcctggaca ggggcctac ttttaagccc tactctggca

FIG. 4H

19/92

```

23161 ctgcctacaa cgcactggcc cccaagggtg cccccaactc gtgcgagtgg gaacaaaatg
23221 aaactgcaca agtggatgct caagaacttg acgaagagga gaatgaagcc aatgaagctc
23281 aggcgcgcga acaggaaaca gctaaagaaa cccatgtata tgcccaggct ccaactgtccg
23341 gaataaaaaa aactaaagaa ggtctacaaa taggaactgc cgagcccaac gtacagctgg
23401 ccggcaaaaga aattttcgca gacaaaactt tccaactcga accacaagta ggagaattct
23461 aatggaacga agcggatgcc acagcagctg gtggaagggt gtggaagggt accacaagta
23521 tgaaccctcg ctatggctca tacgtagac ttggaagtc caacaaatc gcaattttt
23581 tggttgaaca aatggtaaaa aacaatatac actcatcttt cttaaaacc taaaattggg
23641 caaatgccac aatgaagtt caagcaatgc tattacaaca gccacagtaa tgatttgtac
23701 taacatcgga aactccagat caagcaatgc gcttggacaa caactcaatc taattttttt
23761 ccaagtcat gcttggacaa gttgtcatgt gttgtagatt tgcaagacag aaacacagag
23821 acaattttat tgggtctcatg tattacaaca gccacagtaa catgggtgtc ctgtcttacc
23881 aggcacgcga gttgaacgct gttgtagatt ggcgcagaaa caagatactt ttcaattggt aatcaagctg
23941 agctttttgt tgattcaatt gtcaagaata ttgagaacca tgggaactgag gatgagtgtc
24001 ttgacagcta tgatccagat ggtggaattg ggtattactga cacttttcaa gctgttaaaa
24061 caaattattg ctcttctctt caagccaata ctactgggca aagaatttca cacttttttt
24121 caactgtctg taacgggggac ggaataaact tttgccatgga aatlaactct aatggcaaac
24181 aacgcataga aatagggggtg tccaattatg cgtctactct gccagacaag ctaaaataca
24241 ttgtgagaaa ttctctttac tctgacaacc ccaacacctt cgaactcatg aacaagcagc
24301 accccaacca tgtgaaata gactgctaca ttaaccttgg ggcgcgtctg tctctggact
24361 tgggtctgct cgtttaaacc tttaaccaac acgcaatgct gggcctcggt tacgcctcca
24421 acatggacaa cgtttaaacc tttaaccaac acgcaatgct gggcctcggt tacgcctcca
24481 tgttgttggg aaacggccgc taccgtctctg ttcaacttca ggtgccccaa aagttttttt
24541 caattaaaaa cctctctctc ctgcacggct catabacata tgaattgaa ctacaggaa
24601 atgttaacat ggttctgca gactctctgc agactctctg gaaacgacct tagagttgac gggcgtacga
24661 ttaagtttga cagcatttgt ctttaaccga ccttcttccc catggcccac aacacggcct
24721 ccacgctgga agcctatgct agaaatgaca ccaacgacca gtctttaaact gactaccttt
24781 ccgcgcgcaa catgctatat ccatatccgc ccaacgacca caactgtccc atctccatcc
24841 catcgcgcaa ctgggcagca ttctgcggtt gggccttcaac acgcttgaag acaaaagaaa
24901 ccccttccct gggatcaggc tacgacctt actacacct cctcggtctc ataccatacc
24961 ttgacggaac cttctatctt aatcacacct ttaagaaggt ggcattact ttgactctt
25021 ctgttagctg gccgggcaac gaccgcctgc ttaactccaa tgaagtttag attaagcgt
25081 cagttgacgg ggagggctat aacgttagtc agtgcaacat gacaaaggac tggttcctag
25141 tgcagatggt ggccaactac aatatgtgct accagggtct ctacattcca gaaagctaca
25201 aagaccgatg tgaactgttc ttcagaaact tccagcccat gagccggcaa gtgtggagc
25261 atactaaata caaagattat cagcaggttg gaattatcca ccagcataac aactcaggct
25321 tegttagctg cctcgctccc accatgcgcg agggcaagc ttaccgcctc atagttccct
25381 accactaat agggcaaac cgggttgata gtattaccca gaaaagattt ctttgcgacc
25441 gcacctgtgt gcgcacccc ttctccagta acttatgtc catgggtgcg ctacagacc
25501 tggggcaaaa ctttctctac gcaaacctcg cccacgcgtc agacatgacc tttaggttg
25561 atccccatgc cgagcccaac cttctttatg tttgttttga agtcttggac gtgttcggtg
25621 tgcaccagcc gcaccggcgc gtcctcgaga ccgtgtacct gcgcagccc ttctcgccag
25681 gcaacggcac aacataaaga agcaagcaac atcaacaaca gctgcgcgca tggctcccg
25741 tgagcaggaa ctgaagaaca ttgtcaaaga ttttgttgtt gggccatatt tttagggcac
25801 ctatgacaa gcttcccaag gctttgtttc cccacaacag ctgcgcgctg ccatagtaa
25861 caccgcccgt cgcgagactg gggcggtaca ctggatggcc ttgctctgga acccgctct
25921 aaaaacatgc taactctttg agccttttgg acttttgac caacgtctca agcaggttta
25981 ccagttttag taagagtac tcttgcgcgt tagcgcctat gccctttccc ccgaccgctg
26041 tataacgctg gaaaagtcca cccaagctgc cgaggggccc aactgcgccc ctgtgcgct
26101 attctgtctg atgtttctcc acgcttttgc caactggccc caaactccca tggatcaaa
26161 cccaccatg aaacttatta cgggggtacc caactctatg ctaacagctc cccaggtaca
26221 gccaccctg cgccgcaacc aggaacagct ctacagctgt ctggagcgc actcgcccta
26281 cttccgcgac cacagtcgcg aaattaggag cgcactttct tttgtgcact tgaanaacat
26341 gtaaaaaata gtacttagga gacactttca atgtttttat tttgacact
26401 tgggttgatt atttaccccc accttgcgcg tctgcgcgtt taaaaatac aagggtttct

```

FIG. 41

20/92

26461 gccgcgcctc gctatgcgcc actggcaggg acacgttgcg atactggtgt ttagtctctc
 26521 acttaaacctc aggcacaaacc atccgcggca gctcgttgaa gttttcactc cacaggctgc
 26581 gcaccatcac caacgcgttt agcaggtcgg gcgcgcgatat cttgaagtgc cagttggggc
 26641 ctccgccttg cgcgcgcgag ttgcgatatca cagggtttaca gcacttgaac actatcagcg
 26701 ccgggttggtg cagcgtggccc agcacgctct tgcgcggagat cagatcccggt tccaggtctc
 26761 ccgctgtgct caggcggaac ggagtcacat ttggtagctg ccttcccaaa aagggtgcac
 26821 gcccaggctt tgagttgcac tgcacccgta gtggcatcag aagggtgacg tgcccagctc
 26881 gggcgttagg atacagcgcc tgcattgaaa ccttgatctg cttaaaagcc acctgagctc
 26941 ttgcgccttc agagaagaac atgcgcgaag acttgccgga aaactgattg gccgcgacgg
 27001 ccgcgtcatg cagcagcac cttgcgtcgg tgttggagat ctgcaccaca ttctggcccc
 27061 accggtcttt cagcatcttg gcccttgtag actgctcctt cagcgcgcgc tgcgcgtttt
 27121 cgctgcctac atccattca atcacgtgct ccttatttat cataatgctc ccgtgtagac
 27181 acttaagctc gcccttgatc tcagcgacgc ggtgcagcca caacgcgcag ccgctgggct
 27241 cgtggtgctt gtagggtacc tctgcaaacg actgcaggtc cgctcgagg aatcgcccca
 27301 tcactgctac aaagggtctt ttgctggtga aggtcagctg caaccgcggg tgctcctcgt
 27361 tttagcagggt cttgcatacg gcgcgcagag cttccacttg gtcaggcgagt agcttgaagt
 27421 ttgcctttag atcgttatcc acgttggtact tgtccatcaa cgcgcgcgca gcttcdatgc
 27481 cctctcccca cgcagacacg atcggcaggg tcagcggggt tatcaccttt cctcactttt
 27541 ccgcttcaact ggaactcttc ttttctcttt gcaatccgat accccggctg actgggtcgt
 27601 cttcatctcag ccgcgcgacc gtgcgccttc cctccttgcc gtgcttgatt agctgtgatt
 27661 ggttgctgtaa acccaccatt tgtagcgcca catcttctct tctctctcgt cttctccacga
 27721 tcacactctgg ggaatggggg cgctcgggct ttgggagagg gcctctctttt agctttttgg
 27781 acgcaatggc caaatccggc gtgcaggtcg atggcccggg gctgggtgtc cgcggcacc
 27841 gcgcattcttg tgacgagtct tcttctctct cggactcgag acgcgcctct agccgctttt
 27901 ttgggggggc cgggggaggc ggccgcggaag gcgacggggg cggagagctc tccactgtgc
 27961 gtggagctcg cgcgcgaccg cgtccgcgtc cgggggttgt ttgcgcgtgc tctctctccc
 28021 gactggccat tctcttctcc tataggcaga aaaagatcat ggagtcagtc gagaaggagg
 28081 acagctcaac cgcctctctt gagttgcgca ccaccgcctc caccgatgac gcccaacggc
 28141 ctaccacttt ccccgctgag gcacccccgc ttggaggagg ggaagtgtat atcgagcagg
 28201 acccaggctt tgtaagcgaa gacgacgaag atcgctcagt accaacagag gataaaaagg
 28261 aagaccaggga cgacgcagag gcaaacgagg aacaagtctg gcggggggac caaaggcatg
 28321 gcgactcaact agatgtggga gacgacgtgc ttgtgaagca tctgcagcgc cagtgcccca
 28381 ttactctgca cgcgttgcaa gagcgacgag atgtgcccc gcgcatagcg gatgtcagc
 28441 ttgcctacga acgcacactg tctctcaccg cgttaccccc caaacgcgca gaiaacggga
 28501 catcgagagg caacccgcgc ctaactcttc accccgtatt tgcggtgcca gaggttgcgt
 28561 ccactatcca catctttttc caaaactgca agataccctc atcctgcgct gcccaaccga
 28621 gccgagcgga caagcagctg gctttgcgcg agggcgctgt catacctgat atcgctctga
 28681 tcgacgaagt gccaaaaatc tttaggggtc ttggacgcga cgagaagcgc gcggcaacgc
 28741 cttctcaaca agaaaacagc taactctgtg gctactgtgg agtgcctggt gaacttgagg
 28801 gtgacaacgc gcgcctagcc gtgctgaaac gcacagctca ggtcaccacc ttgtgcactc
 28861 cggcacttaa cttacccccc aggtttatga gatgcataa gaggcagctg atcgttgccc
 28921 ttgcacgacc cctggagagg gatgcataa ctgcagaaca aaccgagagg ggctaccgcc
 28981 cagtgggga tgagcagctg gcgcgctggc ttgagacgcg cgagcttgcc gacttgaggg
 29041 agcgacgcaa gctaattgat gccgcagtcg ttgttacctg ggagcttgag tgcactgcag
 29101 ggtctctttg tgaccggagg atgcagcgca agctagagg aacgttgaac taacactttc
 29161 gccagggtca cgtgcgcgag gcttgcaaaa ttcccaactg ggagctctgc acctggtct
 29221 ctacacttgg aattttgcac ttggggcaaaa cgtgcttcat tcccgctcca tccagctga
 29281 agggcgaggc gcgcgcgac gcttgcaaaa gcttgcttta cttattttct tgctacacct
 29341 ggcaaacagc catggggctg ttgcagcagt gcttgaggga gcgaacctg aagagctgc
 29401 aagaagctct aaagcaaaac ttgaaggaac cttagcagcg cttcaacgag cgctccggtg
 29461 ccgcgcactt ggcggaacct atcttccccc atgcctgctc taataacccgt caacagggtc
 29521 ttgcagactt caccagtcga agcatgttgc aaaactttag gaactttatc cttagagctt
 29581 caggagactt gccgcgcacc tgctgtgcgc tctctagcgc ctttctgccc actaagtacc
 29641 gtgaatgcgc tccgcgcctt tggggtaact gctaccttct gcagctagcc aactaccttg
 29701 cctaccactc cgacatcatg gaagcagtgta cgcgtgacgg cctactggag tgtcactgtc

FIG. 4J

21/92

```

29761 gctgcaacct atgcaccccg caccgctccc tggctgtcaa ttcacaaactg cttagcgaaa
29821 gtcaaatatt cggtaacctt gagctgcgag gctccctgcc tgacgaaaag tccgcggctc
29881 cgggggtgaa actcactccg gggctgtgga cgtccggtta ccttcgcmaa ttgttaacctg
29941 aggactacga cgcaccaagc agcaagacca atcccgccgc ccaaatgcgc ccaattgcga
30001 agcttaaccg ctgcgttcatt acccaggggc acatccttgg ccaattgcaa gccatttaaca
30061 aagccgcgca agagtttctg ctacgaaggg gacggggggg ttactttggc cccctggcca
30121 cgaggagcca caaccocaat ccccgccgcg gcagcgccta tcagcagccg cgggcccctg
30181 cttcccgaga tggcacccaa aaagaagctg cagctgcgcg cgcgccaccc cagcgagcag
30241 gaggaatact gggacagtca ggcagaggag gttttggac aggaggaggga gatgatgaa
30301 gactgggaca gccatagacga ggaagcttcc gaggccgaag aggtgtcaga cgaacaacgc
30361 tcacctccgg tcgcatctcc ctgcgcggcg cccacagaat cggcaaccgt tcccagcatt
30421 gctacaaact ccgctcctca ggcgcgcgcg gcactgcgcg ttgcgcgacc caaccgtaga
30481 tgggacacca ctggaaaccag ggcgcggtaa tctaagcagc cgcgcgcgtt accccaagag
30541 caacaaacgc gccaaaggct a ccgctcgtgg cgcggtgcaca agaa cgcgat agtgcgttcg
30601 ttgcaagact gtgggggcaa catctccttc cgcgcgcgct ttcttctcta ccatcaaggc
30661 gtggccttcc cccgtaacat cctgcattac tacgctcatc tctacagccc ctactgcacc
30721 ggcgcgacgc gcagcaaacg cagcgggcca gcagaagcaa aggcgacgag atagacagcg
30781 tctgacaaag cccaagaat cccacagcgc ggcagcagca gggaggaggag cactgcgtct
30841 gtcgcccac gaaccogtat cgcaccgcga gcttgaacaa aggatttttc cactctgtta
30901 tgcataatit caacagagca ggggccaaga acaagagctg aaaaataaaa acaggtctct
30961 gtcgctccct acccgcaegt gcctgtatca caaagcgaaa gatcagcttc cggcagcagct
31021 ggaagacgcy gaggctctct tcagcaaat a ctgcgcgctg actcttaagc actagtttcg
31081 cgcctcttct caaatttaag cgcgaaaaat acgtcatctc cagcgcgcaa acccgcgccg
31141 agcacctgtc gtcagcgcca ttatgagcaa ggaatctccc acgcccata tgtggagtta
31201 gcgcgccaca atgggacttg cgcgtggagc tcccaacagc tactcaacc tactcaacta
31261 catgagcgcy ggaccccaca tgatatcccg ggtcaacgga atccgcgccc accgaacgcg
31321 aattctcttc gaacaggcgy cctattaccac cacaccttgt aataacctta atcccgtag
31381 ttggcccgct gccctggtgt accaggaagc tcccgctccc accactgtg tacttccag
31441 agacgcccag gccgaagtgc agatgactaa ctaaggggcg cagcltgccg gcgctttcg
31501 tcacaggggt cggctgcgcc ggcaggggat aactcacctg aaaaacagag ggcgaggtat
31561 tcagctcaac gacgagtcgg tgagctcctc tcttggttct cgtccggagc ggacattta
31621 gatcgccgcy gctggccgct ctctatttac gcccgctcag gcgactctaa cctgcagac
31681 ctgctcctcg gaggccgct cccgaggcat tggaaactta caatttattg aggagtctgt
31741 gccctcggtt tacttcaacc cctttcttgg acctcccgcc cactaccgcy accagtttat
31801 tcccaacttt gacgcggtaa aagactcggc ggcggtctac gactgaatga ccagtgagga
31861 ggcagagcaa ctgcgcctga cacactcga ccactgcgcg cgcacaagt gctttgcgcg
31921 cgtcctccgt gagttttggc accttgaatt gccgaagag catatcgagg gcccgcgcta
31981 cggcgtccgg ctacccacc caggtagagct tacacgtagc ctgattcggg agtttaccaa
32041 gcgcccccct ctagtggagc ggggcggggg tccctgtgtt ctgacgctgg ttgcaactg
32101 tctcaacctt ggattacatc aagatcttat tccatcaac taacaataaa cacacaataa
32161 attacttact taaaatcagt cagcaaatct ttgtccagct tattcagcat cactcctttt
32221 cctcctcccc aactctggtt ttccagcagc cttttagctg cgaactttct ccaaatgtta
32281 aatggagagt caaatctctc atgtttctgt cctccgcac ccaactttct catattgtgt
32341 cagatgaaac ggcgcagacc gttctgaagc accttcaacc ctgtgtacc atatgacagc
32401 gaacccgccc cctcaactgt gcccttctct acctccctt ttgtgtccc aaatgcagc
32461 caagaaagtc ccccgaggc gcttttttgg cgttcttcag aacttttgtt taactcaac
32521 ggcactgctg cgtcaaaaat gggcagcgcc ctgtccctg atcaggcagc caactctaca
32581 tcaaatacaa tcactgtttc tcaaccgcta aaaaaacaaa agtccaatat aactttggaa
32641 acatccgcgc cctttacagt cagctcagcg gccttaacca tggccaacac ttgccttttg
32701 gtggtctctg acaacactct taccatgcaa tcaacagcac cgtcaaccgt gcagactca
32761 aaacttagca ttgctaccaa agagccactt acaggttag atgaaaactt ggcctcgag
32821 acatcagccc cctctctcgc cactgataac aacgcctcta ctactcgtc ctacctctct
32881 ctactactgt caaatggtag tctggtgtgt acctatgaaa accacttta caaacaat
32941 ggaacacttg ggcctaaaaa tggcggtctt ttgcagtg gcaaccgact acatgacta
33001 acactaggta ctggcaggg ggttgcagtt catacaact ttgtacatac aaaagttaca

```

FIG. 4K

22/92

```

33061 ggcgaatag ggtttgatac atctgggaac atggaactta aaactggaga tggcctctat
33121 gtggatagcg ccggtcctaa ccaaaaacta catatttaac taatatccac aaaagccctt
33181 ccttttgaca acacgcgaat aacaattaac gctggaaaag ggttggaaat tgaaacagac
33241 tctctaaacg gaatcccat aaaaacaaaa attggatcag gcatacaca taataccaat
33301 ggagctatgg ttgcaaaact tgggaacagg cctcagtttg acagctccgg agccataaca
33361 attgggcagca taacaatga cagacttact ctctgggaca caccagacc atccccaaat
33421 tgcagaattg cttcagataa agactgcaag ctaactctgg cgtcaacaaa atgtggcagt
33481 caaattttgg gcaactgttc agctttggca gtatcaggta atatggcctc catcaatgga
33541 actctaagca gtgtaaaact ggttcttaga ttgtatgaca acggagtgtc tatgtcaaat
33601 tcatcactgg acaaacagta ttgggaacttt agaaaacgggg actccactaa cggtaacca
33661 tacacttatg ctgttggggt tatggcaaac ctaaaagctt acccaaaaac tcaaatgaaa
33721 actgcacaaa gtaatatgtg tagccaggtg tatcttaatg gtgacaagtc taaacatgtg
33781 cattttacta ttacgtctaa ttggaacagat gaaaccaacc aagtaagcaa atactcaata
33841 tcattcagtt ggtcctggaa cagtggacaa tacactaatg acaaatgtgc caccattccc
33901 tataccttct cctacattgc ccaggataaa agaactgtga acctgttgca tgttatgttt
33961 caacgtgttt atttttcaat tgcagaaaa ttcaagtcac ttttcattca gtagtatagc
34021 cccaccacca catagcttat actaatcacc gtaccttaat caaactcaca gaaccttagt
34081 attcaaacgt ccacctccct cccaacacac agagtacaca gtcttttccc ccggctggcg
34141 cttaaacagc atcatatcat ggttaacaga catatttcta ggtgttatat tccacacggt
34201 ctctcttga gccaacgcct catcagtgat gtttaataac tccccgggca cctcgtctga
34261 ttctatgtcg ctgtccagct gctgagccac aggtcgttgt ccaacttggg gtctgctaac
34321 gggcgcgcaa gagaagtcg acgcctacat gggggtagag tcataatgtg gtaacagat
34381 agggcggtgg tgcgtcagca gcgcgcgaat aaactgtctg cgcgcgcgtc ccgtctgca
34441 ggaatacaac atggcagtggt tctctcagc gatgatctgc acgcgcgcga gcaaaaggcg
34501 ccttgctctc cgggcacagc aggcacacct gatctcaact aagtcagcac agtaactgca
34561 gcacagtacc acaattatgt taaaaatccc acagtgcagg cgcgtgtatc caaagctcat
34621 ggcggggacc acagaaacca cgtggccatc ataccacaag ataccttttt ggcattgtgt
34681 accctctata aacacgtggt acataaaact tacccttttt ccatggcgca attcaccac
34741 ctcccggtac catataaaac cgtctatgca ctgcagggaa cgggagctgg tccataacca
34801 gctggccaaa acctgcccgc aacctatggt aacctatggt caggattaca agctctctcc cggtcagaac
34861 gtggagagcc caggactcgt tacactttct cagcgttaat cccacactgc agggaaagac
34921 acaacacagc cacacgtgca tacactttct cagcgttaat cccacactgc agggaaagac
34981 catatccagg ggaacaaccc attcctgaat agtgtttacat tccggcgaga cgcgtgatct
35041 tgcacagttaa ctcacgttgt gcaattgtcaa agtgtttacat tccggcgaga cgcgtgatct
35101 ctccagtatg gtacgcgcgg ttctgtcttc aaaaggaggt agacgatccc tactgtacgg
35161 agtgcgcgca gacaaccgag atcgtgttgg tcgtagtgtc atgcacaaatg gaacgcggca
35221 cgtagtcata ttctctgaag caaaacagggt tgcggcgctg atcaaacagat gctcgtctcc
35281 ggtctcgcgc cttagatcgc tctgtgtagt aattgttagta tatcaccctc ctcaaaagcat
35341 ccaagcgcgc cctggctctg ggtttctatg gtttctatgt atgcgcgcgt atgcgcgcgt
35401 catccaccac cgcagaataa gccacaccca gccaacctac acattcgttc tgcgagtcat
35461 acaaggagag agcgggaaga cgtggaagaa ccatgttttt tttttttatc caaaagatta
35521 tccaaaacct caaaatgaag atctattaag tgaacgcgct cccctccggt gccgtgtgta
35581 aactctacag caaagaaca gataatggca ttgttaagat gttgcaaat gttgtccaaa
35641 aggcacaacg ccttcacgtc caagtggagc taaagggtaa acccttcagg gtgaatctcc
35701 tctataaaca ttccagcacc ttcaacatgt cccaataatc tctcatctcg ccaactctcg
35761 aatataatct taagcaaatc cogaatatta agtcgggcca ttgttaaaat ctgtcccaga
35821 gcgcctccca ccttcagcct caagcagcga atcatcattg tttgtcagtg aaaaaattca cgtctctcac
35881 agacctgtat aagattcaaa agcggaaact taacaaaaat acccgcatcc cctagtgtcc
35941 ttgcagagcc cagctgaaca taactcgtga ttgttcgacg gaccagcgcg gccactgtcc
36001 gcgcaggaac catgacaaaa gaacccacac tgattatgac acgcatactc ggaagtatgc
36061 taaccagcgt agccccgatg taagcttgtt cgtatggcgg cgtataaaa tgcaaggtgc
36121 tgcctaaaaa atcaggcaaa gccctcgcga aaaaagaaag cacatcgtag tcatgtcat
36181 gcagataaag caggttaagc tccggaacca ccaagcaaaa agacaccatt tttctctcat
36241 acatgtctgc ggtttctcgc ataaacacaa aataaaataa caaaaaaaca tttaaacatt
36301 agaagcctgt cttacaacag gaaaaaacac ccttataagc ataaagcaga ctacggccat

```

FIG. 4L

23/92

```

36361 gccggcgtga cegtaaaaaa actgggtcacc gtgattaaaa agcaccaccg acagctcctc
36421 ggtcatgtcc gaagtcataa tgtaagactc ggtaaacaca tcagggtgat tcacatcggt
36481 cagtgcataa aagcgaccga aatagcccg ggaatacat acccgaggc gttagagaca
36541 cattacagcc ccctaggag gtataacaaa attaatagga gagaaaaaca cataaacacc
36601 tgaanaacc tcctgcctag gcaaaatagc accctccgc tcagaacaa catcacagcg
36661 tccacagcg gcagccataa cagtcagct taccagtaa aaagaaacc tattaaaaa
36721 acaccactcg acacggcacc agctcaatca gtcacagtgt aaaaaagggc caagtgcaga
36781 gcgagtatat ataggactaa aaatgacgt aacgggttaa gtccacaaa aacaccaga
36841 aaaccgcacg cgaacctacg cccagaaacg aaagccaaa aaccacaa ttcctcaat
36901 cgtcacttcc gttttccac gttacgtcac ttccatttt aagaaaact caattccaa
36961 cacatacaag ttactccgc ctaaaacct cgtcaccgc ccggtccca cgtcccgcg
37021 caggtcacaa actccacccc ctcattatca tattggcttc aatccaaaat aaggtatatt
37081 attgatgatg

```

FIG. 4M

24/92

```

10          30          50
ATGGCGCCCATCACGGCCTACTCCCAACAGACGCGGGCCTACTTGGTTGCATCATCATC
-----+-----+-----+-----+-----+-----+-----+
MetAlaProIleThrAlaTyrSerGlnGlnThrArgGlyLeuLeuGlyCysIleIleThr
          10          20

          70          90          110
AGCCTTACAGGCCGGGACAAGAACCAGGTCGAGGAGAGGTTACGGTGGTTTCCACCGCA
-----+-----+-----+-----+-----+-----+
SerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGlnValValSerThrAla
          30          40

          130          150          170
ACACAATCCTTCCTGGCGACCTGCGTCAACGGCGTGTTGGACCGTTTACCATGGTGCT
-----+-----+-----+-----+-----+-----+
ThrGlnSerPheLeuAlaThrCysValAsnGlyValCysTrpThrValTyrHisGlyAla
          50          60

          190          210          230
GGCTCAAAGACCTTAGCCGGCCCAAGGGGCCAATCACCAGATGTACACTAATGTGSAC
-----+-----+-----+-----+-----+-----+
GlySerLysThrLeuAlaGlyProLysGlyProIleThrGlnMetTyrThrAsnValAsp
          70          80

          250          270          290
CAGGACCTCGTCGGCTGGCAGGCGCCCCCGGGCGCGTTCCTTGACACCATGCACCTGT
-----+-----+-----+-----+-----+-----+
GlnAspLeuValGlyTrpGlnAlaProProGlyAlaArgSerLeuThrProCysThrCys
          90          100

          310          330          350
GGCAGCTCAGACCTTTACTTGGTCACGAGACATGCTGACGTCATTCGGTGCGCGCGGG
-----+-----+-----+-----+-----+-----+
GlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIleProValArgArgArg
          110          120

          370          390          410
GGCGACAGTAGGGGGAGCCTGCTCTCCCCAGGCCTGTCTCTACTTTGAAGGGCTCTTCG
-----+-----+-----+-----+-----+-----+
GlyAspSerArgGlySerLeuLeuSerProArgProValSerTyrLeuLysGlySerSer
          130          140

```

FIG. 5A

25/92

```

      430              450              470
GGTGGTCCACTGCTCTGCCCTTCGGGGCAGCGTGTGGGCATCTTCGGGGCTGCCGTATGC
-----+-----+-----+
GlyGlyProLeuLeuCysProSerGlyHisAlaValGlyIlePheArgAlaAlaValCys
      150              160

      490              510              530
ACCCGGGGGGTTCGGAAGCGGTGGACTTTGTGCCCGTAGAGTCCATGGAACTACTATG
-----+-----+-----+
ThrArgGlyValAlaLysAlaValAspPheValProValGluSerMetGluThrThrMet
      170              180

      550              570              590
CGGTCTCCGGTCTTCACGGACAACCTCATCCCCCGGCCGCTACCGCAGTCATTTCAGTG
-----+-----+-----+
ArgSerProValPheThrAspAsnSerSerProProAlaValProGlnSerPheGlnVal
      190              200

      610              630              650
GCCACCTACACGCTCCCACTGGCAGCGGCAAGAGTACTAAAGTGCCGGCTGCATATGCA
-----+-----+-----+
AlaHisLeuHisAlaProThrGlySerGlyLysSerThrLysValProAlaAlaTyrAla
      210              220

      670              690              710
GCCCCAAGGGTACAAGGTGCTCGTCCTCAATCCGTCCGTTGCCGTACCTTAGGGTTTGGG
-----+-----+-----+
AlaGlnGlyTyrLysValLeuValLeuAsnProSerValAlaAlaThrLeuGlyPheGly
      230              240

      730              750              770
GCGTATATGTCTAAGGCACACGGTATTGACCCCAACATCAGAACTGGGGTAAGGACCATT
-----+-----+-----+
AlaTyrMetSerLysAlaHisGlyIleAspProAsnIleArgThrGlyValArgThrIle
      250              260

      790              810              830
ACCACAGGCGCCCCGTCACATACTCTACCTATGGCAAGTTTCTTGCCGATGGTGGTGC
-----+-----+-----+
ThrThrGlyAlaProValThrTyrSerThrTyrGlyLysPheLeuAlaAspGlyGlyCys
      270              280

```

FIG. 5B

26/92

```

      850              870              890
TCTGGGGGCGCTTATGACATCATATATGTGATGAGTGCCATTCAACTGACTCGACTACA
-----+-----+-----+-----+-----+-----+-----+
SerGlyGlyAlaTyrAspIleIleIleCysAspGluCysHisSerThrAspSerThrThr
      290              300

      910              930              950
ATCTTGGGCATCGGCACAGTCCTGGACCAAGCGGAGACGGCTGGAGCGCGCTTGTCGTG
-----+-----+-----+-----+-----+-----+
IleLeuGlyIleGlyThrValLeuAspGlnAlaGluThrAlaGlyAlaArgLeuValVal
      310              320

      970              990              1010
CTCGCCACCGCTACGCCCTCCGGGATCGGTACCCGTGCCACACCCAAACATCGAGGAGGTG
-----+-----+-----+-----+-----+-----+
LeuAlaThrAlaThrProProGlySerValThrValProHisProAsnIleGluGluVal
      330              340

      1030             1050             1070
GCCCTGTCTAATACTGGAGAGATCCCCTTCTATGGCAAAGCCATCCCCATTGAAGCCATC
-----+-----+-----+-----+-----+-----+
AlaLeuSerAsnThrGlyGluIleProPheTyrGlyLysAlaIleProIleGluAlaIle
      350              360

      1090             1110             1130
AGGGGGGAAGGCATCTCATTTTCTGTCTATCCAAGAAGAAGTGCACGAGCTCGCCGCA
-----+-----+-----+-----+-----+-----+
ArgGlyGlyArgHisLeuIlePheCysHisSerLysLysLysCysAspGluLeuAlaAla
      370              380

      1150             1170             1190
AAGCTGTCAAGCCTCGGAATCAACGCTGTGGCGTATTACCGGGGCTCGATGTGTCGTC
-----+-----+-----+-----+-----+-----+
LysLeuSerGlyLeuGlyIleAsnAlaValAlaTyrTyrArgGlyLeuAspValSerVal
      390              400

      1210             1230             1250
ATACCAACTATCGGAGACGTCGTTGTGCTGGCAACAGCGCTCTGATGACGGGCTATACG
-----+-----+-----+-----+-----+-----+
IleProThrIleGlyAspValValValAlaThrAspAlaLeuMetThrGlyTyrThr
      410              420

```

FIG. 5C

27/92

```

1270          1290          1310
GGCGACTTTGACTCAGTGATCGACTGTAACACATGTGTCACCCAGACAGTCGACTTCAGC
-----+-----+-----+-----+-----+-----+-----+-----+
GlyAspPheAspSerValIleAspCysAsnThrCysValThrGlnThrValAspPheSer
430          440

1330          1350          1370
TTGGATCCACCTTCACCATTGAGACGACGACCGTGCCCTCAAGACGCAGTGTGCGGCTCG
-----+-----+-----+-----+-----+-----+-----+-----+
LeuAspProThrPheThrIleGluThrThrThrValProGlnAspAlaValSerArgSer
450          460

1390          1410          1430
CAGCGCGCGGGTAGGACTGGCAGGGGTAGGAGAGGCATCTACAGGTTGTGACTCCGGA
-----+-----+-----+-----+-----+-----+-----+-----+
GlnArgArgGlyArgThrGlyArgGlyArgArgGlyIleTyrArgPheValThrProGly
470          480

1450          1470          1490
GAACGGCCCTCGGGCATGTTCGATTCTCTCGGTCTGTGTGAGTGCTATGACGCGGGCTGT
-----+-----+-----+-----+-----+-----+-----+-----+
GluArgProSerGlyMetPheAspSerSerValLeuCysGluCysTyrAspAlaGlyCys
490          500

1510          1530          1550
GCTTGGTACGAGCTCACCCCCGCCGAGACCTCGGTTAGGTTGCGGGCTACCTGAACACA
-----+-----+-----+-----+-----+-----+-----+-----+
AlaTrpTyrGluLeuThrProAlaGluThrSerValArgLeuArgAlaTyrLeuAsnThr
510          520

1570          1590          1610
CCAGGGTTGCCCGTTTGCCAGGACCACCTGGAGTTCTGGGAGAGTGCTTCACAGGCCTC
-----+-----+-----+-----+-----+-----+-----+-----+
ProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluSerValPheThrGlyLeu
530          540

1630          1650          1670
ACCCACATAGATGCACACTCTTGTGCCAGACCAAGCAGGCAGGAGACAACCTCCCTAC
-----+-----+-----+-----+-----+-----+-----+-----+
ThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnAlaGlyAspAsnPheProTyr
550          560

```

FIG. 5D

28/92

```

1690          1710          1730
CTGGTAGCATACCAAGCCACGGTGTGCGCCAGGGCTCAGGCCCCACCTCCATCATGGGAT
-----+-----+-----+-----+-----+-----+
LeuValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaProProProSerTrpAsp
          570          580

1750          1770          1790
CAAATGTGGAGTGTCTCATACGGCTGAAACCTACGCTGCACGGGCCAACACCCCTGCTG
-----+-----+-----+-----+-----+
GlnMetTrpLysCysLeuIleArgLeuLysProThrLeuHisGlyProThrProLeuLeu
          590          600

1810          1830          1850
TACAGGCTGGGAGCCGTCCAAATGAGGTCAACCTCACCCACCCCATACCAAAATACATC
-----+-----+-----+-----+-----+
TyrArgLeuGlyAlaValGlnAsnGluValThrLeuThrHisProIleThrLysTyrIle
          610          620

1870          1890          1910
ATGGCATGCATGTCGGCTGACCTGGAGGTCGTCACTAGCACCTGGGTGCTGGTGGGCGGA
-----+-----+-----+-----+-----+
MetAlaCysMetSerAlaAspLeuGluValValThrSerThrTrpValLeuValGlyGly
          630          640

1930          1950          1970
GTCCTTGCAGCTCTGGCGCGTATTGCCTGACAACAGGCAGTGTGGTCATTGTGGGTAGG
-----+-----+-----+-----+-----+
ValLeuAlaAlaLeuAlaAlaTyrCysLeuThrThrGlySerValValIleValGlyArg
          650          660

1990          2010          2030
ATTATCTTGTCCGGGAGGCCGCTATTGTTCCCGACAGGGAGTTTCTTACCAGGAGTTC
-----+-----+-----+-----+-----+
IleIleLeuSerGlyArgProAlaIleValProAspArgGluPheLeuTyrGlnGluPhe
          670          680

2050          2070          2090
GATGAAATGGAAGAGTGCGCTCGCACCTCCCTTACATCGAGCAGGGAATGCAGCTCGCC
-----+-----+-----+-----+-----+
AspGluMetGluGluCysAlaSerHisLeuProTyrIleGluGlnGlyMetGlnLeuAla
          690          700

```

FIG. 5E

[illegible]

FIG. 5F

30/92

```

2530          2550          2570
TTCAAGTTCATGAGCGGCGAGATGCCCTCCACCGAGGACCTGGTCAATCTACTTCTCTGCC
-----+-----+-----+-----+-----+-----+-----+-----+
PheLysValMetSerGlyGluMetProSerThrGluAspLeuValAsnLeuLeuProAla
850          860

2590          2610          2630
ATCCTCTCTCTCTGGCGCCCTGGTCGTCGGGCTCGTGTGTGCAGCACTACTGCGTCGACAC
-----+-----+-----+-----+-----+-----+-----+-----+
IleLeuSerProGlyAlaLeuValValGlyValValCysAlaAlaIleLeuArgArgHis
870          880

2650          2670          2690
GTGGGTCCGGAGAGGGGGCTGTGCAGTGGATGAACCGGCTGATAGCGTTTCGCCCTCGCGG
-----+-----+-----+-----+-----+-----+-----+-----+
ValGlyProGlyGluGlyAlaValGlnTrpMetAsnArgLeuIleAlaPheAlaSerArg
890          900

2710          2730          2750
GGTAATCATGTTTCCCCACGCACTATGTGCCCTGAGAGCGACGCCGAGCGCGTGTATTACT
-----+-----+-----+-----+-----+-----+-----+-----+
GlyAsnHisValSerProThrHisTyrValProGluSerAspAlaAlaAlaArgValThr
910          920

2770          2790          2810
CAGATCCTCTCCAGCCTTACCATCACTCAGCTGCTGAAAAGGCTCCACAGTGGATTAAT
-----+-----+-----+-----+-----+-----+-----+-----+
GlnIleLeuSerSerLeuThrIleThrGlnLeuLeuLysArgLeuHisGlnTrpIleAsn
930          940

2830          2850          2870
GAAGACTGCTCCACACCGTGTTCGGCTCGTGGCTAAGGGATGTTTGGGACTGGATATGC
-----+-----+-----+-----+-----+-----+-----+-----+
GluAspCysSerThrProCysSerGlySerTrpLeuArgAspValTrpAspTrpIleCys
950          960

2890          2910          2930
ACGGTGTGACTGACTTCAAGACCTGGCTCCAGTCCAAGCTCCTGCCGAGCTACCGGGA
-----+-----+-----+-----+-----+-----+-----+-----+
ThrValLeuThrAspPheLysThrTrpLeuGlnSerLysLeuLeuProGlnLeuProGly
970          980

```

FIG. 5G

31/92

```

2950                2970                2990
GTCCCTTTTTCCTCGTGCCAACGCGGGTACAAAGGGAGTCTGGCGGGGAGACGGCATCATG
-----+-----+-----+-----+-----+-----+-----+-----+
ValProPhePheSerCysGlnArgGlyTyrLysGlyValTrpArgGlyAspGlyIleMet
                      990                                1000

3010                3030                3050
CAAACCACCTGCCCATGTGGAGCACAGATCACCGGACATGTCAAAAACGGTTCCATGAGG
-----+-----+-----+-----+-----+-----+-----+
GlnThrThrCysProCysGlyAlaGlnIleThrGlyHisValLysAsnGlySerMetArg
                      1010                                1020

3070                3090                3110
ATCGTCGGGCCCTAAGACCTGCAGCAACACGTGGCATGGAACATTCGCCATCAACGCATAC
-----+-----+-----+-----+-----+-----+-----+
IleValGlyProLysThrCysSerAsnThrTrpHisGlyThrPheProIleAsnAlaTyr
                      1030                                1040

3130                3150                3170
ACCACGGGCCCCGTCACACCCCTCTCCAGCGCCAAACTATTCTAGGGCGCTGTGGCGGGTG
-----+-----+-----+-----+-----+-----+-----+
ThrThrGlyProCysThrProSerProAlaProAsnTyrSerArgAlaLeuTrpArgVal
                      1050                                1060

3190                3210                3230
GCCGCTGAGGAGTACGTGGAGGTACGCGGGTGGGGGATTTCCACTACGTGACGGGCATG
-----+-----+-----+-----+-----+-----+-----+
AlaAlaGluGluTyrValGluValThrArgValGlyAspPheHisTyrValThrGlyMet
                      1070                                1080

3250                3270                3290
ACCACTGACAACGTAAGTGCCCATGCCAGGTTCCGGCTCCTGAATTCTTCACGGAGGTG
-----+-----+-----+-----+-----+-----+-----+
ThrThrAspAsnValLysCysProCysGlnValProAlaProGluPhePheThrGluVal
                      1090                                1100

3310                3330                3350
GACGGAGTGCGGTTGCACAGGTACGCTCCGGCGTGCAGGCCTCTCCTACGGGAGGAGTT
-----+-----+-----+-----+-----+-----+-----+
AspGlyValArgLeuHisArgTyrAlaProAlaCysArgProLeuLeuArgGluGluVal
                      1110                                1120

```

FIG. 5H

32/92

```

3370          3390          3410
ACATTCAGGTCGGGCTCAACCAATACCTGGTGGGTCACAGCTACCATGCGAGCCCGAA
-----+-----+-----+-----+-----+-----+
ThrPheGlnValGlyLeuAsnGlnTyrLeuValGlySerGlnLeuProCysGluProGlu
1130          1140

3430          3450          3470
CCGGATGTAGCAGTGTCTCACTTCCATGCTACCGACCCCTCCCATCACAGCAGAAACG
-----+-----+-----+-----+-----+-----+
ProAspValAlaValLeuThrSerMetLeuThrAspProSerHisIleThrAlaGluThr
1150          1160

3490          3510          3530
GCTAAGCGTAGGTGGCCAGGGGTCCTCCCCCTCTTGCCAGCTCTTCACTAGCCAG
-----+-----+-----+-----+-----+-----+
AlaLysArgArgLeuAlaArgGlySerProProSerLeuAlaSerSerAlaSerGln
1170          1180

3550          3570          3590
TTGTCTGCGCCTTCCTTGAAGGCGACATGCATACCCACCATGTCTCTCCGACGCTGAC
-----+-----+-----+-----+-----+-----+
LeuSerAlaProSerLeuLysAlaThrCysThrThrHisHisValSerProAspAlaAsp
1190          1200

3610          3630          3650
CTCATCGAGGCCAACCTCCTGTGGCGGAGAGATGGGCGGGAACATCACCCGCGTGGAG
-----+-----+-----+-----+-----+-----+
LeuIleGluAlaAsnLeuLeuTrpArgGlnGluMetGlyGlyAsnIleThrArgValGlu
1210          1220

3670          3690          3710
TCGGAGAACAAAGGTGGTAGTCTTGGACTCTPTTCGACCCGCTTCGAGCGGAGGAGTAC
-----+-----+-----+-----+-----+-----+
SerGluAsnLysValValValLeuAspSerPheAspProLeuArgAlaGluGluAspGlu
1230          1240

3730          3750          3770
AGGGAAGTATCCGTTCGGCGGAGATCCTGCGGAAATCCAAGAAGTTCCTCCGCGCAGTAC
-----+-----+-----+-----+-----+-----+
ArgGluValSerValProAlaGluIleLeuArgLysSerLysLysPheProAlaAlaMet
1250          1260

```

FIG. 51

33/92

```

      3790              3810              3830
CCCATCTGGGCGCGCCCGATTACAACCTCCACTGTTAGAGTCCTGGAAGGACCCGGAC
-----+-----+-----+-----+-----+-----+-----+-----+
ProIleTrpAlaArgProAspTyrAsnProProLeuLeuGluSerTrpLysAspProAsp
                        1270                      1280

      3850              3870              3890
TACGTCCTCCCGGTGGTGCACGGGTGCCCGTTGCCACCTATCAAGGCCCTCCAATACCA
-----+-----+-----+-----+-----+-----+-----+-----+
TyrValProProValValHisGlyCysProLeuProProIleLysAlaProProIlePro
                        1290                      1300

      3910              3930              3950
CCTCCACGGAGAAAGAGGACGGTTGTCTCTAACAGAGTCCTCCGTGTCTTCGTCCTTAGCG
-----+-----+-----+-----+-----+-----+-----+-----+
ProProArgArgLysArgThrValValLeuThrGluSerSerValSerSerAlaLeuAla
                        1310                      1320

      3970              3990              4010
GAGCTCGCTACTAAGACCTTCGGCAGCTCCGAATCATCGGCCGTGCACAGCGGCACGGCG
-----+-----+-----+-----+-----+-----+-----+-----+
GluLeuAlaThrLysThrPheGlySerSerGluSerSerAlaValAspSerGlyThrAla
                        1330                      1340

      4030              4050              4070
ACCGCCCTTCCTGACCAGGCCTCCGACGACGGTGACAAAGGATCCGACGTTGAGTCGTAC
-----+-----+-----+-----+-----+-----+-----+-----+
ThrAlaLeuProAspGlnAlaSerAspAspGlyAspLysGlySerAspValGluSerTyr
                        1350                      1360

      4090              4110              4130
TCCTCCATGCCCCCCTTGAGGGGGAACCGGGGACCCCGATCTCAGTGACGGGTCTTGG
-----+-----+-----+-----+-----+-----+-----+-----+
SerSerMetProProLeuGluGlyGluProGlyAspProAspLeuSerAspGlySerTrp
                        1370                      1380

      4150              4170              4190
TCTACCGTGAGCGAGGAAGCTAGTGAGGATGTCGTCTGCTGCTCAATGTCTACACATGG
-----+-----+-----+-----+-----+-----+-----+-----+
SerThrValSerGluGluAlaSerGluAspValValCysCysSerMetSerTyrThrTrp
                        1390                      1400

```

FIG. 5J

34/92

```

      4210              4230              4250
ACAGGCGCCTTGATCAGCCATGCGCTGCGGAGGAAAGCAAGCTGCCCATCAACGCGTTG
-----+-----+-----+-----+-----+-----+-----+-----+
ThrGlyAlaLeuIleThrProCysAlaAlaGluSerLysLeuProIleAsnAlaLeu
                        1410              1420

      4270              4290              4310
AGCAACTCTTTGCTGCGCCACCATAACATGGTTTATGCCACAACATCTCGCAGCGCAGGC
-----+-----+-----+-----+-----+-----+-----+-----+
SerAsnSerLeuLeuArgHisHisAsnMetValTyrAlaThrThrSerArgSerAlaGly
                        1430              1440

      4330              4350              4370
CTGCGGCAGAGAAGGTCACCTTTGACAGATGCAAGTCCTGGACGACCACTACCGGGAC
-----+-----+-----+-----+-----+-----+-----+-----+
LeuArgGlnLysLysValThrPheAspArgLeuGlnValLeuAspAspHisTyrArgAsp
                        1450              1460

      4390              4410              4430
GTGCTCAAGGAGATGAAGCGAAGGCGTCCACAGTTAAGGCTAAACTCCTATCCGTAGAG
-----+-----+-----+-----+-----+-----+-----+-----+
ValLeuLysGluMetLysAlaLysAlaSerThrValLysAlaLysLeuLeuSerValGlu
                        1470              1480

      4450              4470              4490
GAAGCCTGCAAGCTGACGCCCCACATTGCGCCAAATCCAAGTTTGGCTATGGGGCAAAG
-----+-----+-----+-----+-----+-----+-----+-----+
GluAlaCysLysLeuThrProProHisSerAlaLysSerLysPheGlyTyrGlyAlaLys
                        1490              1500

      4510              4530              4550
GAGGTCGGAACCTATCCAGCAAGGCGTTAACCACATCCACTCCGTGTGGAAGGACTTG
-----+-----+-----+-----+-----+-----+-----+-----+
AspValArgAsnLeuSerSerLysAlaValAsnHisIleHisSerValTrpLysAspLeu
                        1510              1520

      4570              4590              4610
CTGGAAGACACTGTGACACCAATTGACACCACCATCATGGCAAAAAATGAGGTTTCTGT
-----+-----+-----+-----+-----+-----+-----+-----+
LeuGluAspThrValThrProIleAspThrThrIleMetAlaLysAsnGluValPheCys
                        1530              1540

```

FIG. 5K

35/92

```

      4630              4650              4670
GTCCAACCAGAGAAAGAGGCCGTAAGCCAGCCCGCCTTATCGTATTCCCAGATCTGGGA
-----+-----+-----+-----+-----+-----+-----+
ValGlnProGluLysGlyGlyArgLysProAlaArgLeuIleValPheProAspLeuGly
      1550              1560

      4690              4710              4730
GTCCGTGTATGCGAGAAAGATGGCCCTCTATGATGTGGTCGCCACCTTCCTCAGGTCGTG
-----+-----+-----+-----+-----+-----+-----+
ValArgValCysGluLysMetAlaLeuTyrAspValValSerThrLeuProGlnValVal
      1570              1580

      4750              4770              4790
ATGGGCTCCTCATACGGATTCCAGTACTCTCCTGGCAGCGAGTCGAGTTCTCTGGTGAAT
-----+-----+-----+-----+-----+-----+-----+
MetGlySerSerTyrGlyPheGlnTyrSerProGlyGlnArgValGluPheLeuValAsn
      1590              1600

      4810              4830              4850
ACCTGGAAATCAAAGAAAAACCCATGGGCTTTTCATATGACACTCGCTGTTTCGACTCA
-----+-----+-----+-----+-----+-----+-----+
ThrTrpLysSerLysLysAsnProMetGlyPheSerTyrAspThrArgCysPheAspSer
      1610              1620

      4870              4890              4910
ACGGTACCCGAGAACGACATCCGTGTTGAGGAGTCAATTTACCAATGTTGTGACTTGCC
-----+-----+-----+-----+-----+-----+-----+
ThrValThrGluAsnAspIleArgValGluGluSerIleTyrGlnCysCysAspLeuAla
      1630              1640

      4930              4950              4970
CCCGAAGCCAGACAGGCCATAAAATCGCTCACAGACGGCTTTATATCGGGGCTCTCTG
-----+-----+-----+-----+-----+-----+-----+
ProGluAlaArgGlnAlaIleLysSerLeuThrGluArgLeuTyrIleGlyGlyProLeu
      1650              1660

      4990              5010              5030
ACTAATTCAAAAGGGCAGAAGTCGCGGTATTATCGCCGGTGCCGCGCAGCGGCGTGCTGACG
-----+-----+-----+-----+-----+-----+-----+
ThrAsnSerLysGlyGlnAsnCysGlyTyrArgArgCysArgAlaSerGlyValLeuThr
      1670              1680

```

FIG. 5L

5050	5070	5090
ACTAGCTGCGGTAAACACCCCTCACATGTTACTTGAAGGCCTCTGCAGCCTGTGCAGCTGCGG		
ThrSerCysGlyAsnThrLeuThrCysTyrLeuLysAlaSerAlaAlaCysArgAlaAla		
	1690	1700
5110	5130	5150
AAGCTCAGGACTGCACGATGCTCGTGAAACGGAGACACCTTGTCTTATCTGTGAAAGC		
LysLeuGlnAspCysThrMetLeuValAsnGlyAspAspLeuValValIleCysGluSer		
	1710	1720
5170	5190	5210
GCGGGAACCCAAGAGGACGCGGCCGAGCCTACGAGTCTTCACGGAGGCTATGACTAGGTAC		
AlaGlyThrGlnGluAspAlaAlaSerLeuArgValPheThrGluAlaMetThrArgTyr		
	1730	1740
5230	5250	5270
TCTGCCCCCCCGGGGACC GCGCCCAACCAAGTAACGACTTGGAGCTGATAACATCATGT		
SerAlaProProGlyAspProProGlnProGluTyrAspLeuGluLeuIleThrSerCys		
	1750	1760
5290	5310	5330
TCCTCCAATGTGTGGTGC GCCCACGATGCATCAGGCAAAGGGTG TACTACCTCACCCGT		
SerSerAsnValSerValAlaHisAspAlaSerGlyLysArgValTyrTyrLeuThrArg		
	1770	1780
5350	5370	5390
GATCCCAACCAACCCCTCGCACGGGCTGC GTGGGAAACAGCTAGACA CACTCCAGTTAAC		
AspProThrThrProLeuAlaArgAlaAlaTrpGluThrAlaArgHisThrProValAsn		
	1790	1800
5410	5430	5450
TCCTGGCTAGGCAACATTATCATGTATGCGCCCACTTTGTGGGCAAGGATGATTCTGATG		
SerTrpLeuGlyAsnIleIleMetTyrAlaProThrLeuTrpAlaArgMetIleLeuMet		
	1810	1820

FIG. 5M

37/92

```

5470          5490          5510
ACTCACTTCTCTCCATCCTTCTAGCACAGGAGCAACTTGAAAAGCCCTGGACTGCCAG
-----+-----+-----+-----+-----+-----+-----+
ThrHisPhePheSerIleLeuLeuAlaGlnGlnLeuGluLysAlaLeuAspCysGln
1830          1840

5530          5550          5570
ATCTACGGGCGCTGTACTCCATTGAGCCACTTGACCTACCTCAGATCATTGAACGACTC
-----+-----+-----+-----+-----+-----+-----+
IleTyrGlyAlaCysTyrSerIleGluProLeuAspLeuProGlnIleIleGluArgLeu
1850          1860

5590          5610          5630
CATGCCTTAGCGCATTTCCTACTCCATAGTTACTCTCCAGGTGAGATCAATAGGGTGCT
-----+-----+-----+-----+-----+-----+-----+
HisGlyLeuSerAlaPheSerLeuHisSerTyrSerProGlyGluIleAsnArgValAla
1870          1880

5650          5670          5690
TCATGCCTCAGGAACTTGGGGTACCACCTTGCGAGTCTGGAGACATCGGGCCAGGAGC
-----+-----+-----+-----+-----+-----+-----+
SerCysLeuArgLysLeuGlyValProProLeuArgValTrpArgHisArgAlaArgSer
1890          1900

5710          5730          5750
GTCCGCGCTAGGCTACTGTCCCAGGGGGGAGGGCCGCACTTGTGGCAAGTACCTCTTC
-----+-----+-----+-----+-----+-----+-----+
ValArgAlaArgLeuLeuSerGlnGlyGlyArgAlaAlaThrCysGlyLysTyrLeuPhe
1910          1920

5770          5790          5810
AACTGGGCAGTGAAGACCAAACTCAAATCACTCCAAATCCCGGTGCGTCCCAGCTGGAC
-----+-----+-----+-----+-----+-----+-----+
AsnTrpAlaValLysThrLysLeuLysLeuThrProIleProAlaAlaSerGlnLeuAsp
1930          1940

5830          5850          5870
TTGTCCGGCTGGTTCTGCTGGTTACAGCGGGGAGACATATACAGCCTGTCTCGT
-----+-----+-----+-----+-----+-----+-----+
LeuSerGlyTrpPheValAlaGlyTyrSerGlyGlyAspIleTyrHisSerLeuSerArg
1950          1960

```

FIG. 5N

38/92

```

5890          5910          5930
GCCCGACCCGCTGGTTCATGCTGTGCCTACTCCTACTTCTGTAGGGGTAGGCATCTAC
-----+-----+-----+-----+-----+
AlaArgProArgTrpPheMetLeuCysLeuLeuLeuLeuSerValGlyValGlyIleTyr
          1970          1980

```

```

5950 5955
CTGCTCCCAACCGA (SEQ. ID. NO. 5)
-----+-----
LeuLeuProAsnArg (SEQ. ID. NO. 6)
          1985

```

FIG. 50

39/92

1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG
 51 GAGACGGTCA CAGCTTGCTT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG
 101 TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACATAG
 151 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA
 201 CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA
 251 TTGCATACGT TGTATCCATA TCATAATATG TACATTATAT TTGGCTCATG
 301 TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTATAGT
 351 AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT
 401 ACATAACTTA CGGTAAATGG CCCGCTGGC TGACCGCCCA ACGACCCCG
 451 CCCATTGACG TCAATAATGA CGTATGTTCC CATAGTAACG CCAATAGGGA
 501 CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC TGCCCACTTG
 551 GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
 601 TGACGGTAAA TGGCCCCCTT GGCATTATGC CCACTACATG ACGTTATGGG
 651 ACTTTCCTAT TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG
 701 GTGATGCGGT TTTGGCAGTA CATCAATGGG CGTGATAGC GGTTTGACTC
 751 ACGGGGATTT CCAAGTCTCC ACCCCATTGA CGTCAATGGG AGTTTGTTTT
 801 GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCTTAACAA CTCGCCCAAC
 851 TTGACGCAAA TGGCGGGTAG GCGGTGACGG TGGGAGGTCT ATATAAGCAG
 901 AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT
 951 TTGACCTCCA TAGAAGACAC CGGACCGGAT CCAGCCTCCG CGGCCGGGAA
 1001 CGGTGCATTG GAACGCGGAT TCCCCTGGC AAGAGTGACG TAAGTACCGC
 1051 CTATAGACTC TATAGGCACA CCCCTTTGGC TCTTATGCAT GCTATACTGT
 1101 TTTTGGCTTG GGGCTTATAC ACCCCCGCTT CCTTATGCTA TAGGTGATGG
 1151 TATAGCTTAG CCTATAGGTG TGGGTTATTG ACCATTATTG ACCACTCCCC
 1201 TATTGGTGAC GATACTTTCC ATTACTAATC CATAACATGG CTCTTTGCCA
 1251 CAACTATCTC TATTGGCTAT ATGCCAATAC TCTGTCCCTT AGAGACTGAC
 1301 ACGGACTCTG TATTTTACAA GATGGGGTCC CCATTATATTA TTTACAAATT
 1351 CACATATACA ACAACGCGGT CCCCGTGGC CGCAGTTTAT ATTAAACATA
 1401 GCGTGGGATC TCCACGCGAA TCTCGGGTAC GTGTTCCGGA CATGGGCTCT
 1451 TCTCCGGTAG CGCGGAGCT TCCACATCCG AGCCCTGGTC CCATGCCTCC
 1501 AGCGGCTCAT GGTTCGCTCG CAGCTCCTTG CTCTTAACAG TGGAGGCCAG
 1551 ACTTAGGCAC AGCAACAATG CCACCACCAC CAGTGTGCCG CACAAGGCCG
 1601 TGGCGGTAGG GTATGTCTCT GAAATAGAG GTGGAGATTG GGCCTGCACG
 1651 GCTGACGCGA ATGGAAGACT TAAGCGAGCG GCAGAAGAAG ATGCAGGCAG
 1701 CTGAGTTGTT GTATTCTGAT AAGAGTCAGA GGTAACTCCC GTTGCGGTGC
 1751 TGTTAACGGT GGAGGGCAGT GTAGTCTGAG CAGTACTCGT TGCTGCCCGG
 1801 CGCGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTTCCAT
 1851 GGGTCTTTTC TGCAGTCACC GTCTTATGAT CTAGGTACCA GATATCAGAA
 1901 TTCAGTCGAC AGCGGCCCGG ATCTGCTGTG CCTTCTAGTT GCCAGCCATC
 1951 TGTGTGTTTC CCTCCCCCGG TGCCCTTCTT GACCCCTGAA GTGCCACTC
 2001 CCACGTGCTT TTTCTAATAA AATGAGGAAA TTGCATCGCA TTGCTCTGAT
 2051 AGGTGTCAAT CTATTCTGGG GGGTGGGGTG GGGCAGGACA GCAAGGGGGA

FIG. 6A

40/92

2101 GGATTGGGAA GACAATAGCA GGCATGCTGG GGATGCGGTG GGCTCTATGG
 2151 CCGCTGCGGC CAGGTGCTGA AGAATTGACC CGGTTCCTCC TGGGCCAGAA
 2201 AGAAGCAGGC ACATCCCTTT CTCTGTGACA CACCTGTCC ACGCCCTGG
 2251 TTCTTAGATC CAGCCCACT CATAGGACAC TCATAGCTCA GGAGGGCTCC
 2301 GCCTTCAATC CCACCCGCTA AAGTACTTGG AGCGGTCTCT CCTCCCTCA
 2351 TCAGCCCAAC AAACCAAAAC TAGCCTCCAA GAGTGGGAAG AAATTAAAGC
 2401 AAGATAGGCT ATTAAGTGCA GAGGGAGAGA AAATGCCCTC AACATGTGAG
 2451 GAAGTAATGA GAGAAATCAT AGAATTCTCT CGCTTCCTC GCTCACTGAC
 2501 TCGCTGCGCT CGTCTGTTGG GCTGCGGCGA GCGGTATCAG CTCACCTCAA
 2551 GGCCTGTAAT CGTTATTCCT CAGAATCAGG GGATAACGCA GGAAGAGACA
 2601 TGTGAGCAAA AGGCCAGCAA AAGGCCAGGA ACCGTAAAAA GGCCTGCTTG
 2651 CTGGCGTTTT TCCATAGGCT CCGCCCCCTT GACGAGCATC AAAAAAATCG
 2701 ACGCTCAAGT CAGAGGTGCG GAAACCCGAC AGGACTATAA AGATACCAGG
 2751 CGTTTCCCCC TGGAAAGCTC CTGCTGCGCT CTCTGTCTCC GACCTGCGCG
 2801 CTTACCGGAT ACCTGTCCGC CTTTCTCCCT TCGGGAAGCG TGGCGCTTTC
 2851 TCATAGCTCA CGCTGTAGGT ATCTCAGTTC GGTGTAGGTC GTTCGCTCCA
 2901 AGCTGGGCTG TGTCCACGAA CCCCCGTTT AGCCCGACCG CTGCGCTCTA
 2951 TCCGGTAACT ATCGTCTTGA GTCCAAACCG GTAAGACACG ACTTATCGCC
 3001 ACTGGCAGCA GCCACTGTGA ACAGGATTAG CAGAGCGAGG TATGTAGGCG
 3051 GTGCTACAGA GTTCTTGAAG TGGTGGCCTA ACTACGGCTA CACTAGAAGA
 3101 ACAGTATTGG GTATCTGCGC TCTGCTGAAG CCAATTACCT TCGGAAAAAG
 3151 AGTTGTAGTC TCTTGATCCG GCAACAAAC CACCGCTGGT AGCGGTGGTT
 3201 TTTTGTGTTG CAAGCAGCAG ATTACGCGCA GAAAAAAGG ATCTCAAGAA
 3251 GATCCTTTGA TCTTTCTTAC GGGGTCTGAC GCTCAGTGGA ACGAAAATC
 3301 ACGTTAAGGG ATTTTGSTCA TGAGATTATC AAAAAGGATC TTCACCTAGA
 3351 TCCTTTTAAA TTAATAATGA AGTTTAAAT CAATCTAAAG TATATATGAG
 3401 TAAACTTGGT CTGACAGTTA CCAATGCTTA ATCAGTGAGG CACCTATCTC
 3451 AGCGATCTGT CTATTTCGTT CATCCATAGT TGCCTGACTC GGGGGGGGGG
 3501 GGCCTGTAGG TCTGCCTCGT GAAGAAGGTG TTGCTGACTC ATACCGGCC
 3551 TGAATCGCCC CATCATCCAG CCAAGAAAGT AGGGAGCCAC GGTGTATGAG
 3601 AGCTTTGTGT TAGGTGGACC AGTTGGTGAT TTGTAACTTT TGCTTTGCCA
 3651 CGGAACGCTG TGCCTTGTGG GGAAGATGCG TGATCTGATC CTTCAACTCA
 3701 GCAAAAGTTC GATTATTATCA ACAAAGCCGC CGTCCCGTCA AGTCAGCGTA
 3751 ATGCTCTGCC AGTGTATCAA CCAATTAAAC AATTCTGATT AGAAAACTC
 3801 ATCGAGCATC AAATGAAACT GCAATTTATC CATATCAGGA TTATCAATAC
 3851 CATATTTTTG AAAAAAGCGT TTCTGTAATG AAGGAGAAAA CTCACCGAGG
 3901 CAGTTCCATA GGATGGCAAG ATCCTGGTAT CGGTCTGCGA TTCCGACTCG
 3951 TCCAACATCA ATACAACCTA TTAATTTCCC CTGCTCAAAA ATAAGGTTAT
 4001 CAAGTGAGAA ATCACCATGA GTGACGACTG AATCCGCTGA GAATGGCAAA
 4051 AGCTTATGCA TTTCTTTCCA GACTTGTTC AAGGCCAGC CATTACGCTC
 4101 GTCAATCAAA TCACTCGCAT CAACCAAAAC GTTATTCAAT CGTGATTGCG
 4151 CCTGAGCGAG ACGAAATACG CGATCGCTGT TAAAAGGACA ATTACAAACA

FIG. 6B

41/92

4201 GGAATCGAAT GCAACCGGCG CAGGAACACT GCCAGCGCAT CAACAATATT
4251 TTCACCTGAA TCAGGATATT CTCTAATAC CTGGAATGCT GTTTTCCCGG
4301 GGATCGCAGT GGTGAGTAAC CATGCATCAT CAGGAGTACG GATAAAATGC
4351 TTGATGGTCG GAAGAGGCAT AAATTCCGTC AGCCAGTTTA GTCTGACCAT
4401 CTCATCTGTA ACATCATTTGG CAACGCTACC TTTGCCATGT TTCAGAAAACA
4451 ACTCTGGCGC ATCGGGCTTC CCATACAATC GATAGATTGT CGCACCTGAT
4501 TGCCCGACAT TATCGCGAGC CCATTTATAC CCATATAAAT CAGCATCCAT
4551 GTTGGAATTT AATCGCGGCC TCGAGCAAGA CGTTTCCCGT TGAATATGGC
4601 TCATAACACC CTTTGTATTA CTGTTTATGT AAGCAGACAG TTTTATTGTT
4651 CATGATGATA TATTTTATC TTGTGCAATG TAACATCAGA GATTTTGAGA
4701 CACAACGTGG CTTTCCCCC CCCCCATTA TTGAAGCATT TATCAGGGTT
4751 ATTGTCTCAT GAGCGGATAC ATATTTGAAT GTATTTAGAA AAATAAACAA
4801 ATAGGGGTTT CGCGCACATT TCCCCGAAA GTGCCACCTG ACGTCTAAGA
4851 AACCATTATT ATCATGACAT TAACCTATAA AATAGGCGT ATCACGAGGC
4901 CCTTTCGTC

FIG. 6C

42/92

1 CATCATCAAT AATATACCTT ATTTTGGATT GAAGCCAATA TGATAATGAG GGGGTGGAGT
 61 TTGTGACGTG GCGCGGGCG TGGGAACGGG GCGGGTGACG TAGTAGTGTG GCGGAAGTGT
 121 GATGTTGTAA GTGTGCGGA ACACATGTAA GCGCCGGATG TGCTAAAAGT GACGTTTTTG
 181 GTGTGCGCGG GTGTACACGG GAAGTGACAA TTTTCGCGCG GTTTTAGGGC GATGTTGTAG
 241 TAAATTTGGG CGTAACCAAG TAATATTGG CCAATTTTCG GGGAAAAGT AATAAGAGGA
 301 AGTGAATCT GAATAATCT GTGTTACTCA TAGCGCGTAA TATTGTCTA GGGCCGCGG
 361 GACTTTGACC GTTTACGTGG AGACTCGCCC AGGTGTTTT CTCAGGTGT TTCCGCGTTC
 421 CGGGTCAAAG TTGGCGTTTT ATTATTATAG TCAGCTGACG CGCAGTGTAT TTATACCCG
 481 TGAGTTCCTC AAGAGGCCAC TCCTGAGTGC CAGCGAGTAG AGTTTTCTCC TCCGAGCCGC
 541 TCCGACACCG GGACTGAAAA TGAGACATAT TATCTGCCAC GGAGGTGTTA TTACCGAAGA
 601 AATGGCCCGC AGTCTTTTGG ACCACGTGAT CGAAGAGGTA CTGGCTGATA ATCTTCACC
 661 TCCTAGCCAT TTTGAACCA CTAACCTTCA CGAAGTGTAT GATTTAGAGC TGACGGCCCC
 721 CGAAGATCCC AACGAGGAGG CGGTTTCGCA GATTTTTCCC GAGTCTGTAA TGTGGCGGT
 781 CGAGGAAGGG ATTGACTTAT TCACTTTTCC GCGCGCGCCC GGTCTCCGG AGCGCGCTCA
 841 CCTTCCCGCG CAGCCCGAGC AGCCCGGAGC GAGAGCCTTG GGTCCGGTT CTATGCCAAA
 901 CCTGTGCGCG GAGGTGATCG ATCTTACCTG CCACGAGGCT GGCCTTCCAC CCAAGTACGA
 961 CGAGGATGAA GAGGGTGAGG AGTTTGTGTT AGATTATGTT GAGCACCCCG GGCACGGTTG
 1021 CAGGCTCTGT CATTATCACG GGAGGAATAC GGGGGACCCA GATATTATGT GTTCGCTTTG
 1081 CTATATGAGG ACCTGTGSCA TGTITGTCTA CAGTAAGTGA AAAATTATGG CCAAGTGGTG
 1141 ATAGAGTGGT GGGTTTGGTG TGGTAATTTT TTTTITAAAT TTTACAGTTT TGTGGTTTAA
 1201 AGAATTTTGT ATTGTGATTT TTTAAAGGT CCTGTGCTG AACCTGAGCC TGAGCCGAGG
 1261 CCAGAACCAG AGCCTGCAAG ACCTACCCCG CGTCTCTAAT TGGTGCTGCT TATCTCAGGA
 1321 CGCCCGACAT CACCTGTGTC TAGAGAAATG AATAGTAGTA CGGATAGCTG TGACTCCGGT
 1381 CCTTCTAACA CACCTCCTGA GATACACCG GTGTCTCCCG TGTGCCCAT TAAACAGTT
 1441 GCGGTGAGAG TTGGTGGGCG TCGCCAGGCT GTGGAATGTA TCGAGGACIT GCTTAACGAG
 1501 TCTGGGCAAC CTTTGGACTT GAGCTGTAAA CGCCCCAGGC CATAAGGTGT AAACCTGTGA
 1561 TTGCGTGTGT GGTAAACGCC TTTGTTGCT GAATGAGTTG ATGTAAGTTT AATAAAGGT
 1621 GAGATAATGT TTAACITGCA TGGCGTGTAA AATGGGCGG GGCCTAAAGG GTATATAATG
 1681 CGCCGTGGGC TAATCTTGGT TACATCTGAC CTCATGGAGG CTTGGGAGTG TTTGGAAGAT
 1741 TTTCTGCTG TGCCTAAGTT GCTGGAACAG AGCTCTAACA GTACCTCTTG GTTTTGGAGG
 1801 TTTCTGTGGG GCTCTCCCA GGCATAAGTT GTCTGCAGAA TTAAGGAGGA TTAAGATGG
 1861 GAATTTGAAG AGCTTTTGAA ATCTGTGGT GAGCTGTTTG ATTCTTTGAA TCTGGGTAC
 1921 CAGCGCGCTT TCCAAGAGAA GGTCAACAG ACTTTGGATT TTTCCACACC GGGAGCGGCT
 1981 GCGGCTGCTG TTGCTTTTTT GAGTTTTATA AAGGATAAAT GGAGCGAAGA AACCACTCTG
 2041 AGCGGGGGGT ACCTGCTGGA TTTTCTGGCC ATGCATCTGT GGAGAGCGGT GTGAGACAC
 2101 AAGAACTGCC TGCTACTGTT GTCTTCCGTC CGCCCGGCAA TAATACCGAC GAGCGCGGCT
 2161 CAGCAGGAGG AAGCCAGGCG GCGCGGCGG CAGGAGCAGA GCCCATGGAA CCCGAGAGCC
 2221 GGCCTGGACC CTCGGGAATG AATGTTGTAT AGGTGGCTGA ACTGTTTCCA GAACGTGAGC
 2281 GCATTTTAA CATTAAACGAG GATGGGCGAG GGTAAAGGG GGTAAAGAG GAGCGGGGG
 2341 CTTCTGAGGC TACAGAGGAG GCTAGGAATC TAACTTTTAG CTTAATGACC AGACACCGTC
 2401 CTAGTGTGT TACTTTTCAG CAGATTAAAG ATAATTGCGC TAATGAGCTT GATCTGCTGG
 2461 CGCAGAAGTA TTCCATAGAG CAGCTGACCA CTTACTGGCT CGAGCCAGGG GATGATTTTG

FIG. 7A

43/92

2521 AGGAGGCTAT TAGGGTATAT GCAAAGGTGG CACTTAGGCC AGATTGCAAG TACAAGATTA
 2581 GCAAACTTGT AAATATCAGG AATTGTTGCT ACATTTCCTGG GAACGGGGCC GAGGTGGAGA
 2641 TAGATACGGA GGATAGGGTG GCTTTAGATG GTAGCATGAT AAATATGTGG CCGGGGGTGC
 2701 TTGGCATGGA CGGGGTGGTT ATTATGAATG TGAGGTTTAC TGGTCCCAAT TTTAGCGGTA
 2761 CGGTTTTCTT GGCCAATACC AATCTTATCC TACACGGTGT AAGCTTCTAT GGGTTTAACA
 2821 ATACCTGTGT GGAAGCCTGG ACCGATGTAA GGGTTCGGGG CTGTGCCCTT TACTGCTGCT
 2881 GGAAGGGGGT GGTGTGTCGC CCCAAAAGCA GGGCTTCAAT TAAGAAATGC CTGTTTGAAA
 2941 GGTGTACCTT GGGTATCCTG TCTGAGGGTA ACTCCAGGGT GCGCCACAAT GTGGCCTCCG
 3001 ACTGTGGTTG CTTTATGCTA GTGAAAAGCG TGGCTGTGAT TAAGCATATC ATGGTGTGTG
 3061 GCAACTGCGA GGACAGGGCC TCTCAGATGC TGACCTGCTC GGACGGCAAC TGTCACTTGC
 3121 TGAAGACCAT TCACGTAGCC AGCCACTCTC GCAAGGCCTG GCCAGTGTTC GAGCACAAAC
 3181 TACTGACCCG CTGTTCTTTG CATTTCGGTA ACAGGAGGGG GGTGTTCTTA CCTTACCAAT
 3241 GCAATTTTGG TCACACTAAG ATATTGCTTG AGCCCGAGAG CATGTCCAAG GTGAACCTGA
 3301 ACGGGGTGTT TGACATGACC ATGAAGATCT GGAAGGTGCT GAGGTACGAT GAGACCCGCA
 3361 CACTGCGCAG ACCCTGCGAG TGTGGCGGTA AACATATTAAG GAACCAGCCT GTGTGCTGG
 3421 ATGTGACCGA GGAGCTGAGG CCCGATCACT TGGTGTCTGC CTCGACCCGC GCTGAGTTTG
 3481 GCTCTAGCGA TGAAGATACA GATTGAGGTA CTGAAATGTG TGGGCGTGGC TTAAGGTTGG
 3541 GAAAGAAATAT ATAAGGTGGG GGTCTCATGT AGTTTGTAT CTGTTTTCGA GCAAGCGCGG
 3601 CCATGAGCGC CAACTCGITT GATGGAAGCA TTGTGAGCTC ATATTTGACA ACGCGCATGC
 3661 CCCATGGGCG CGGGGTGCGT CAGAATGTGA TGGGCTCCAG CATTGATGGT CGCCCCGTCC
 3721 TGCCCGCAAA CTCTACTACC TTGACCTTACG AGACCGTGTC TGGAAACGCG TTGGAGACTG
 3781 CAGCCTCCGC CGCCGCTTCA GCCGCTGCAG CCACCGCCCG CGGGATTGTG ACTGACTTTG
 3841 CTTTCTGTAG CCCGCTTGCA AGCAGTGCGAG CTTCCTCGTT ATCCGCCCCG GATGACAAAT
 3901 TGAAGGCTCT TTTGGCACAA TTGGATTCTT TGACCCGGGA ACTTAATGTC GTTTCTCAGC
 3961 AGCTGTGTGA TCTGCGCCAG CAGGTTTCTG CCTGAAGGC TTCTTCCCTC CCCAATGCGG
 4021 TTTAAACAT AAATAAAAC CAGACTCTGT TTGGATTGG ATCAAGCAAG TGTCITGCTG
 4081 TCTTTATTTA GGGGTTTTGC GCGCGCGGTA GCGCCGGGAC CAGCGGCTCT GTCGTGTTAG
 4141 GGTCTCTGTG ATTTTTCCTA GGACGTGGTA AAGGTGACTT TGGATGTTCA GATACATGGG
 4201 CATAGGCCCG TCTCTGGGGT GGAGGTAGCA CCACTGCAGA GCTTCATGCT GCGGGGTGCT
 4261 GTTGTAGATG ATCCAGTCGT AGCAGGAGCG CTGGGCGTGG TGCCCTAAAA TGTCTTTCAG
 4321 TAGCAAGCTG ATTGCCAGGG GCAGGCCCTT GGTGTAAAGT TTTTCAAAAGC GGTAAAGCTG
 4381 GGAATGGGTG ATACGTGGGG ATATGAGATG CATCTTGGAC TGTATTTTTA GGTGGCTAT
 4441 GTTCCAGGCC ATATCCCTCC GGGGATTCAAT GTTGTGCAGA ACCACCAGCA CAGTGATATC
 4501 GGTGCACTTG GGAATTTTGT CATGTAGCTT AGAAGGAAT GCGTGGAGA ACTTGGAGAC
 4561 GCCTTTGTGA CCTCCAAGAT TTTCTCATGA TTCGTCCATA ATGATGCAAA TGCGCCACG
 4621 GCGCGCGGCC TGGGCGAAGA TATTTCTGGG ATCACTAACG TCATAGTTGT GTTCCAGGAT
 4681 GAGATCGTCA TAGGCCATTT TTACAAAGCG CGGGCGGAGG GTGCCAGACT GCGGTATATAT
 4741 GGTTCCTATCC GGCCAGGGG CGTAGTTACC CTCACAGATT TGCATTTCCT ACGCTTTGAG
 4801 TTCAGATGGG GGGATCATGT CTACCTGCGG GCGCATGAAG AAAACCGTTT CCGGGGTAGG
 4861 GGAGATCAGC TTGGGAAGAA GCAGGTTCTT AAGCAGCTGC GACTTACCCG AGCGGTGGG
 4921 CCGTAAATC ACACCTATTA CCGGCTGCAR CTGGTAGTTA AGAGAGCTGC AGCTGCGCTC
 4981 ATCCCTGAGC AGGGGGGCCA CTTCGTTAAG CATGTCCCTG ACTTGATGT TTTCCCTGAC

FIG. 7B

44/92

5041 CAAATCCGCC AGAAGGCGCT CGCCGCCAG CGATAGCAGT TCTTGCAAGG AAGCAAAGTT
 5101 TTTCAACGGT TTGAGGCCGT CCGCGCTAGG CATGCTTTTG AGCGTTTGAC CAAGCAGTTC
 5161 CAGGCGGTCC CACAGCTCGG TCACGTGCTC TACGGCATCT CGATCCAGCA TATCTCCTCG
 5221 TTTCCGCGGT TGGGCGGGCT TTCGTGTAC GGCAGTAGTC GGTGCTCGTC CAGACGGGCC
 5281 AGGCTCATGT CTTTCCACGG GCGCAGGGTC CTGCTCAGCG TAGTCTGGGT CACGGTGAAG
 5341 GGGTGCCTCT CGGGTTGCGC GCTGGCCAGG GTGCGCTTGA GGTGTGCTCT GCTGTGTGCTG
 5401 AAGCGCTGCC GGTCTTCGCC CTGCGCGTCG GCCAGGTAGC ATTTGACCAT GGTGTCTATG
 5461 TCCAGCCCTC CCGCGGCGTG GCGCTTGCGC GCGAGCTTGC CCTTGGAGGA GGCGCCGCAC
 5521 GAGGGGACGT GCAGACTTTT AAGGGCGTAG AGCTTGGGCG CGAGAAATAC CGATTCCGGG
 5581 GAGTAGGCAT CCGCGCCGCA GCGCCCGCAG ACGGTCTCGC ATTCCACGAG CCAAGTGAGC
 5641 TCTGGCCGTT CGGGGTCAAA AACCAAGGTT CCCCCTATGT TTTTGATGCG TTTCTTACCT
 5701 CTGTTTCCCA TGAGCCGGTG TCCACGCTCG GTGACGAAAA GGTGTGTCCG GTCCCCGTAT
 5761 ACAGACTTGA GAGCCTCTGC CTCGAGCGGT GTTCCGCGGT CCTCTCTGTA TAGAAACTCG
 5821 GACCACCTCT AGACGAAGGC TCGCGTCCAG GCCAGCACGA AGGAGGCTAA GTGGGAGGGG
 5881 TAGCGGTGCT GTGCCACTAG GGGGTCCACT CGCTCCAGGG TGTGAAGACA CATGTCCGCC
 5941 TCTTCGGCAT CAAGGAAGGT GATTGGTTTA TAGGTGTAGG CCACGTGACC GGGTGTTCCT
 6001 GAAGGGGGGC TATAAAAGGG GGTGGGGGCG CGTTCGTCTC CACTCTCTTC CGCATCGCTG
 6061 TCTGCGAGGG CCAGCTGTGT GGGTGAAGTAC TCCTCTCAA AAGCGGGCAT GACTTCTGCG
 6121 CTAGATATGT CAGTTTCTAA AAGCAGGAGG GATTGTATAT TCACCTGGCC CGCGGTGATG
 6181 CCTTTGAGGG TGCCCGCGTC CATCTGGTCA GAAAGACAA TCTTTTGTGT GTCAAGCTTG
 6241 GTGGCAAACG ACCCGTAGAG GCGCTTGGAC AGCAACTTGG CGATGGAGCG CAGGGTTTGG
 6301 TTTTGTCTCG GATCGCGCGC CTCCTTGGCC GCGATGTTTA GCTGCACGTA TTCGCGCGCA
 6361 ACGCACCGCC ATTCTGGAAA GACGGTGGTG CGCTCGTCGG GCACTAGGTG CACGCGCCAA
 6421 CCGCGGTTGT CAGCGGTGAC AAGGTCAAAG CTGGTGGCTA CCTCTCCGCG TAGGCGCTCG
 6481 TTGGTCCAGC AGAGGCGGCC GCCCTTGCGC GAGCAGAATG GCGGTAGTGG GTCTAGCTGC
 6541 GTCTCTCCCG GGGGCTCTGC GTCCACGGTA AAGACCCCGG GCAGCAGGCG CGCGTCAAG
 6601 TAGTCTATCT TGATCTCTTG GCATCTCTAG CCCTGTGTCG ATGCGCGGGC GGCAMCGCG
 6661 CGCTCGTATG GGTGTAGTGG GGGACCCCAT GGCATGGGGT GGGTGAGCCG GGAGGCGTAC
 6721 ATGCGCAAAA TGTCGTAAAC GTAGAGGGGC TCTCTGAGTA TTCCAAGATA TGTAGGGTAG
 6781 CATCTTCCAC GCGGATGCTT GCGCGCACG TAATCGTATA GTTCTGTGCA GGGAGCGGAG
 6841 AGGTGCGGAC CGAGGTTGCT ACGGGCGGGC TGCTCTGCTC GGAAGACTAT CTGCGCTAAG
 6901 ATGGCATGTG AGTTGATGTA TATGTTTGA CGCTGGAAGA CGTTGAAGCT GGCCTCTGTG
 6961 AGACCTTACC GCTACGCAC GAGGAGGCG TAGGAGTCCG GCAGCTTGTG GACGAGCTCG
 7021 GCGGTGACCT GCACGTCTAG GCGCAGTAG TCCAGGGTTT CCTTGATGAT GTCATACTTA
 7081 TCCTGTCCCT TTTTTCCTCA CAGCTCGCGG TTGAGGACAA ACTCTTCCGG GTCTTTCCAG
 7141 TACTCTTGA TCGGAAACCC GTCGGCTTCC AAGCGGTAG AGCCTAGCAT GTAGAACTGG
 7201 TTGACGGGCT GGTAGCGCCA GCATCCCTTT TCTACGGGTA GCGCGTATGC CTGCGCGGCC
 7261 TTCCGAGGCG AGGTGTGGGT GAGCGCAAG GTGTCCCTAA CCATGACTTT GAGGTATCTG
 7321 TATTTGAAGT CAGTGTCTGT GCATCCGCCC TGCTCCGAGA GCAAAAAGTC CGTGGCTTT
 7381 TTGGAACCGG GGTTTGGCAG GCGGAAGGTG ACATCGTTGA AGAGTATCTT TCCCGCGCGA
 7441 GGCATAAAGT TCGCTGTGAT GCGGAAGGCT CCCGCCACCT CGGAACGGTT GTTAATATAC
 7501 TGGGCGCGGA GCACGATCTC GTCAAAGCGG TTGATGTTGT GGCCACAAT GTAAAGTTCC

FIG. 7C

45/92

7561 AAGAAGCGCG GGATGCCCTT GATGGAAGGC AATTTTTTAA GTTCCTCGTA GGTGAGCTCT
 7621 TCAGGGGAGC TGAGCCCGTG CTCTGAAAGG GCCCAGTCTG CAAGATGAGG GTTGGAAGCG
 7681 ACGAATGAGC TCCACAGGTC ACGGGCCATT AGCATTTGCA GGTGGTCGCG AAAGTCCCTA
 7741 AACTGGCGAC CTATGGCCAT TTTTCTGGG GTGATGCAGT AGAAGGTAA GCGGCTCTGT
 7801 TCCACGCGGT CCCATCCAAG GTCCGCGGCT AGGTCTCGCG CGCGGGTCAC TAGAGGCTCA
 7861 TCTCCGCCGA ACTTCATGAC CAGCATGAAG GGCACGAGCT GCTTCCCAA GGGCCCCATC
 7921 CAAGTATAGG TCTCTACATC GTAGGTGACA AAGAGACGCT CGGTCCGNGG ATGCGAGCCG
 7981 ATCGGGAAGA ACTGGATCTC CCGCCACCAG TTGGAGGAGT GGTCTGTGAT GTGGTGAAGA
 8041 TAGAAGTCCC TCGACCGGGC CGAACACTCG TGCTGGCTTT TGTAAAAACG TCGCAGTATC
 8101 TGGCAGCGGT GCACGGGCTG TACATCTGCG ACGAGGTGTA CCTGACGACC GCGCAACAAG
 8161 AAGCAGAGTG GGAATTTGAG CCCCTCGCCT GCGGGGTTTG GCTGTGGTCT TTCTACTTCG
 8221 GCTGCTTGTC CTTGACCCTC TGGCTGCTCG AGGGGAGTTA CGGTGGATCG GACCACCACG
 8281 CCGCGCGAGC CCAAAGTCCA GATGTCCGCG CGCGCGGGTC GGAGCTTGAT GACAACATCG
 8341 CGCAGATGGG AGCTGTCCAT GGTCTGGAGC TCCCGCGCGC TCAGGTTCAGG CGGGAGCTCC
 8401 TCGAGGTTTA CTTCCGATAG CCGGGTCAGG GCGCGGGCTA GGTCCAGGTG ATACCTGTTA
 8461 TCCAGGGGCT GGTGTGTGGC GCGCTCGATG GCTTGCAAGA GGCCGCATCC CCGCGCGCGG
 8521 ACTACGCTAC CGCGCGCGCG GCGGTGGGCG GCGGGGGTGT CCTTGGATGA TCCATCTAAA
 8581 AGCGGTGACG CGGCGGGGCC CCGCGAGGTA GGGGGGGCTC GGGACCCGCG GGGAGCGGGG
 8641 GCAGGGGCAC GTCGGCGCGC CGCGCGGGCA GGAGCTGGTG CTGCGCGCGG AGGTGTCTGG
 8701 CGAACGCGAC GACGCGGCGG TTGATCTCCT GAATCTGGCG CCTCTGCGTG AAGACGACGG
 8761 GCCCGGTGAG CTTGAACCTG AAGAGAGTTT CGACAGAAAT AATTTGCGTG TCGTGTACGG
 8821 CGGCTTGGCG CAAAATCTCC TGCACGTCTC CTGAGTTGTC TTGATAGGCG ATCTCGGCCA
 8881 TGAATCTCTC GATCTCTTCC TCTTGGAGAT CTCGCCGTCC GGTCTGCTCC ACGGTGCGCG
 8941 CGAGGTTCGT GAGATGCGCG GCCATGAGCT GCGAGAAGGC GTTGAGGCTT CCTCTGTTCT
 9001 AGACGCGGCT GTAGACCACG CCCCTTCGCG CATCGCGGGC GCGCATGACC ACCTGCGCGA
 9061 GATTGAGCTC CACGTGCCGG GCGAAGACGG CGTAGTTTTC CAGGCGCTGA AAGAGGTAGT
 9121 TGAGGGTGGT GCGGGTGTGT TCTGCCACGA AGAAGTACAT AACCACGCGC CGCAACGTGG
 9181 ATTCTGTGAT ATCCCCCAAG GCCTCAAGGC GCTCCATGGC CTCGTAGAAG TCCACGCGCA
 9241 AGTTGAAAAA CTGGGAGTTG CGCGCCGACA CGGTAACTC CTCCTCCAGA AGACGGATGA
 9301 GCTCGGCGAC AGTGTGCGCG ACCTCGCGCT CAAAGGCTAC AGGGGCTCTT TCTTCTCTTT
 9361 CAATCTCTCT TCCATAAAGG GCCTCCCTTT CTCTCTCTTC TGCGCGCGGT GGGGGAGGGG
 9421 GGACACGGCG GCGCAGCAGG CGCACCGGGA GCGGGTCGAC AAGCGCTCGC ATCATCTCC
 9481 CGCGGCGACG GCGCATGGTC TCGGTGACGG CGCGGCGGTT CTCGCGGGGG CGCAGTTGGA
 9541 AGACCGCCGCC GCTCATGTCC CGGTATGGG TGTGGCGGGG GCTGCGCGTC GCGAGGGATA
 9601 CGCGCGTAA C GATGCATCTC AACAAATTGTT GTGTAGGTAC TCCGCCACCG AGGACCTCGA
 9661 GCGAGTCCGC ATCGACCCGA TCGGAAAACC TCTCGAGAAA GCGGCTCTAC CAGTCACAGT
 9721 CGCAAGGTAG CGTGAGCACC GTGCGGGGCG GCAGCGGGCG GCGGTCTGGG TTGTTCTTCG
 9781 CGAGGTGCT GCTGATGATC TAATTAAGT AGCGGCTCTT GAGACGGCGG ATGGTCGACA
 9841 GAAGCACCAT GTCTTGGGT CCGGCTGTCT GAATGCGCAG CGGGTCGGCC ATGCCCCAGG
 9901 CTTCTGTTTG ACATCGCGCG AGGTCTTGT AGTAGTCTTG CATGAGCCTT TCTACCGGCA
 9961 CTTCTTCTCT TCTTCTCTCT TGTCTCTGAT CTCTTGCATC TATCGCTGCG CGCGCGGCGG
 10021 AGTTTGGCGG TAGGTGGCGC CCTCTTCTCT CCATGCTGTG GACCCCGAAG CCCCCTCATG

FIG. 7D

46/92

10081 GCTGAAGCAG GCCAGGTCG GCGACAACGC GCTCGGCTAA TATGGCCTGC TGCACCTGGG
 10141 TGAGGGTAGA CTGGAAGTCG TCCATGTCCA CAAAGCGGTG GTATGCGCCG GTGTTGATGG
 10201 TGTAAGTGCA GTTGGCCATA ACGGACCACT TAACGGGCTG GTGACCCGCG TGCAGAGACT
 10261 CGGTGTACTT GAGACGCGAG TAAGCCCTTG AGTCAAGAGC GTAGTCCGTT CAAAGTCGCA
 10321 CCAGGTACTG GTATCCACAC AAAAAGTGGC GCGCGCGCTG GCGGTAGAGG GGCACGGCTA
 10381 GGGTGGCCGG GGCTCCGGGG GCGAGGTCTT CCAACATAAG GCGATGATAT CCGTAGATGT
 10441 ACCTGGACAT CCAGGTGATG CCGGCGGCGG TGGTGGAGGC GCGCGGAAAG TCACGGACGC
 10501 GGTTCACAGT GTTGCGCAGC GGC AAAAAGT GCTCCATGGT CCGGACGCTC TGGCGGCTCA
 10561 GCGCGCGGCA GTGCTTGACG CTCTAGACCG TGCAAAAGGA GAGCCTGTAA GCGGGCACTC
 10621 TTCCGTGTGT TGGTGGATAA ATTCCGAAGG GTATCATGGC GGACGACCGG GGTTCGAACC
 10681 CCGGATCCGG CCGTCCGCGC TGATCCATGC GGTACCGCC CCGGTGTGCA ACCCAGGTGT
 10741 GCGAGCTCAG ACAACGGGGG AGCGCTCCTT TTGGCTTCTT TCCAGGCGCG GCGGATGCTG
 10801 CGCTAGCTTT TTTGGCCACT GCGCGCGCGC GCGCTAAGCG GTTAGGCTGG AAAGCGAAAG
 10861 CATTAAGTGG CTCGCTCCCT GTAGCCGAGG GGTATATTTT CAAGGGTTGA GTGCGGGAC
 10921 CCCCGTTTCG AGTCTCGGGC CGGCGGACT GCGCGCAAGC GGGGTTTGCC TCCCCGTCAT
 10981 GCAAGACCCC GCTTGCAAAAT TCCCTCGGAA ACAGGGACGA GCCCTTTTTC TGCTTTTCCC
 11041 AGATGCATCC GGTGCTCGCG CAGATGCGCC CCCTCTCTCA GCAGCGGCA GAGCAAGAC
 11101 AGCGCGACAG ATCGAGGCGA CCTCTCCCTT CTCTTACCGG GTACAGGAGG GCAACATCCG
 11161 CGCTGACGCG GCGCGCAGAT GGTGATTACG AACCCCGCG GCGCCGACAC CGCACTACTT
 11221 TGGACTTGGA GGAGGGCGAG GGCCTGGCGC GGCATAGGAG GCCCTCTCTT GAGCGACACC
 11281 CAAGGTTGCA GCTGAAGCGT GACACGCGCG AGGCGTACGT GCGCGCGCAG AACCTGTTTT
 11341 GCGACCCGGA GGGAGAGGAG CCCGAGGAGA TCGCGGATCG AAAGTTCCAT GCAGGCGCG
 11401 AGTTGCGGCA TGGCCTGAAC CGCGAGCGGT TGCTGCGCGA GGAGGACTTT GAGCCCGAGC
 11461 CGCGGACCGG GATTAGTCCC GCGCGCGCAC ACGTGGCGGC GCGCGACCTG GTAAACCGCT
 11521 ACGAGCGACG GGTGAACCAAG GAGATTAACT TTCAAAAAGC CTTTAAACAC CACGTGCGCA
 11581 CGCTTGTGGC GCGCGAGGAG GTGGCTATAG GACTGATGCA TCTGTGGGAC TTTGTAAGCG
 11641 CGCTGGAGCA AAACCCAAAT AGCAAGCCGC TCATGGCGCA GCTGTTCCTT ATAGTGACAG
 11701 ACAGCAGGGA CAACGAGGCA TTCAGGATG CGCTGCTAAA CATAGTAGAG CCCGAGGGCC
 11761 GCTGGCTGCT CGATTGATA AACATTCTGC AGAGCATAGT GGTGCAAGAG CGCAGCTTGA
 11821 GCTGGGCTGA CAAGGTGGCC GCCATTAACT ATTCCATGCT CAGCTGTGGC AAGTTTATAG
 11881 CCCGCAAGAT ATACCATACC CCTTACGTTC CCATAGACAA GGAGGTAAG ATCAGAGGGT
 11941 TCTATATGCG CATGGCGCTG AAGGTGCTTA CCTTGAAGCA CGACCTTGGC GTTATTCGCA
 12001 ACAGAGCGAT CCACAGGGCC GTGAGCGTGA GCGCGCGCGC CGAGCTAGC GACGCGGAGC
 12061 TGATGCACAG CCTGCAAAAG GCCCTGGCTG GCACGGGCG GCGCGATAGA GAGGCCGAGT
 12121 CTTACTTTGA GCGGGGCGCT GACCTGCGCT GGGCCCCAAG CCGAGCGCGC CTGGAAGCGAG
 12181 CTGGGGCGCG ACCTGCGCTG CCGTGGCAC CCGCGCGCGC TGGCAAGCTG GCGGCGGTGG
 12241 AGGAATATGA CGAGGACGAT GAGTACGAGC CAGAGGACGG CGAGTACTAA GCGGTGATGT
 12301 TTTGATCAG ATGATGCAAG ACGCAACGGA CCGCGCGGTG CCGGCGCGCC TCGAGAGCCA
 12361 GCGGTCCGGC CTTAACTCCA CCGACGACTG GCGCGAGGTC ATGAGACGCA TCAATGTGCT
 12421 GACTGCGGCG AACCTTGACG CGTTCGGGCA GCGAGCCGAG GCCAACCGCG TCTCCGCAAT
 12481 TCTGGAAGCG GTGGTCCCGG CGCGCGCAA CCCACGCAAC GAGAAGGTGC TGGCGATGCT
 12541 AAACGCGCTG GCCGAAAACA GGGCCATCCG GCCCGATGAG GCGCGGCTGG TCTACGACGC

FIG. 7E

47/92

12601 GCTGCTTCAG CGCGTGGCTC GTTACACAG CAGCAACGTG CAGACCAACC TGGACCGGCT
 12661 GGTGGGGGAT GTGCGCGAGG CCGTGGCGCA GCGTGAGCGC GCGCAGCAGC AGGGCAACCT
 12721 GGGCTCCATG GTTGCACTAA ACGCCTTCCT GAGTACACAG CCCGCCAACG TGCCGCGGGG
 12781 ACAGGAGGAC TACACCAACT TTGTGAGCGC ACTGGCGCTA ATGGTGACTG AGACACCGCA
 12841 AAGTGAGGTG TATCAGTCCG GGGCAGACTA TTTTTCOCAG ACCAGTAGAC AAGGCCTGCA
 12901 GACCGTAAAC CTGAGCCAGG CTTTCAAGAA CTTCGAGGGG CTGTGGGGGG TGCGGGCTCC
 12961 CACAGCGCAC CGCGCGACCG TGCTAGCTT GCTGACGCC AACTCGCGCC TGTTGCTGCT
 13021 GCTAATAGCG CCCTTCACGG ACAGTGGCAG CGTGCTCCGG GACACATACC TAGGTCACTT
 13081 GCTGACACTG TACCGCGAGG CCATAGGTCA GCGCATGTG GACGAGCATA CTTTCCAGGA
 13141 GATTACAAGT GTTAGCCCGG CGCTGGGCGA GGAGGACACG GGCAGCCTGG AGGCAACCTT
 13201 GAAC TACCTG CTGACCAACC GCGCGCAAAA AATCCCTCG TTGCACAGTT TAAACAGCGA
 13261 GGAGGAGCGC ATTTTGGGCT ATGTGACGCA GAGCGTGAGC CTTAACCTGA TGCGCGACGG
 13321 GGTAACGCCG AGCGTGGCGC TGGACATGAC CGCGCGCAAC ATGGAACCGG GCATGTATGC
 13381 CTCACAAACG CCGTTTATCA ATCGCCTAAT GGACTACTTG CATCGCGCGG CGCGCGTGAA
 13441 CCGCGAGTAT TTCAACCAATG CCATCTTGAA CCGCGACTGG CTACCGCCCG CTGGTTCTTA
 13501 CACCGGGGGA TTGAGGTGTC CCGAGGGTAA CGATGGATTG CTCTGGGAGC ACATAGACGA
 13561 CAGCGTGTGT TCCCGCAAC CGCAGACCTT GCTAGAGTTG CAACAACCGG AGCAGCGAGA
 13621 GCGCGCGCTG GGAAGGAAGA GCTTCCGAG GCGAAGCAGC TTGTCCGATC TAGGCGCTGC
 13681 GGGCCCGCGG TCAGATGCTA GTAGCCCAT TCCAAGCTTG ATAGGGTCTC TTACCAAGCAC
 13741 TCGCACCAAC CGCCCGCGCC TGCTGGGCGA GGAGGAGTAC CTAACAACAT CGCTGCTGCA
 13801 GCGCGAGCGC GAAAAGAACC TGCTCCGGC GTTCCCAAC AACGGGATAG AGAGCCTAGT
 13861 GGACAAGATG AGTAGATGGA AGACGTATGC GCAGGAGCAC AGGGATGTGC CCGGCCCGCG
 13921 CCGGCCCAAC CGTCGTCAAA GGCACGACCG TCAGCGGGGT CTGGTGTGGG AGGACGATGA
 13981 CTCGCGAGAC GACAGCAGCG TCTTGGATT TTGGAGGAGT GGCACCCGT TTGCACACT
 14041 TCGCCCCAGG CTGGGGAGAA TGTTTTAAAA AAAGCATGAT GCAAAAATAA AAACCTACCA
 14101 AGGCCATGGC ACGAGCGGTT GGTTCCTCTG TATTCGCCCT AGTATGCGCG CGCGCGCGAT
 14161 GTATGAGGAA GGTCTCTCTC CTTCTACGA GAGCGTGGT AGCGCGCGCG CAGTGGCGCG
 14221 GCGGCTGGGT TCACCTCTCG ATGCTCCCTT GGACCCGCGG TTGCTGCCTC CGCGGTACCT
 14281 GCGGCTTACC GGGGGGAGAA ACAGCATCCG TTACTCTGAG TTGGCACCCC TATTCGACAC
 14341 CACCCGTGTG TACCTTGTGG ACAACAAGTC AACGGATGTG GCATCCCTGA ACTACGAGAA
 14401 CGAGCACAGC AACTTTCTAA CCACGGTCAT TCACAAACAT GACTACAGCC GCGGGGAGGC
 14461 AAGCACACAG ACCATCAATC TTGACGACCG GTCGCACTGG GCGGCGGACC TGAACCACTA
 14521 CCTGCATACC AACTGCCCCA ATGTGAACGA GTTCATGTTT ACCAATAAGT TTAAGGCGCG
 14581 GGTGATGGTG TCGCGCTCGC TTAATAAGGA CAACACAGGT GAGCTGAAT ACAGTGGGT
 14641 GAGATTCACG CTGCCCGAGG GCAACTACT CGAGACCATG ACCATAGACC TTATGAACAA
 14701 CGCGATCGTG GAGCACTACT TGAAGTGGG CAGGCAGAAC GGGTTCTCG AAAGCGACAT
 14761 GCGGGTAAAG TTTGACACCC GCACTTCAG ACTGGGGTTT GACCCAGTCA CTGCTCTTGT
 14821 CATGCTCGG GTATATACAA ACGAAGCCTT CCATCCAGAC ATCATTTTGC TGCGAGGATG
 14881 CGGGGTGGAC TTCACCCACA GCGCGCTGAG CAACTTGTG GGCATCCGCA AGCGGCAACC
 14941 CTTCACGAGG GGCTTTAGGA TCACCTACGA GACTCTGGAG GGTGGTAAAC TTCCGCACT
 15001 GTTGGATGTG GACGCTTACC AGGCAAGCTT GAAAGATGAC ACCGAACAGG GCGGGGTTGG
 15061 CGCAGCGCGC GGCAACACAA GTGGCAGCGG CGCGGAAGAG AACTCCAACG CGGCAGCTGC

FIG. 7F

48/92

15121 GGCAATGCAG CCGGTGGAGG ACATGAACGA TCATGCCATT CGCGGCGACA CCTTTGCCAC
 15181 ACGGGCGGAG GAGAAGCGCG CTGAGGCCGA GGCACGGCCG GAAGCTGCGC CCCC CGCTGC
 15241 GGAGGCTGCA CAACCCGAGG TCGAGAAAGCC TCAGAGAAAG CCGGTGATTA AACCCCTGAC
 15301 AGAGGACAGC AAGAAACGCA GTTACAACCT AATAAGCAAT GACAGCACCT TCACCCAGTA
 15361 CGCGAGCTGG TACCTTGCAAT ACAACTACGAG CGACCCCTCAG GCCGGGATCC GCTCATGGAC
 15421 CCTGCTTTGCT ACTCCTGACG TAACCTGCGG CTCGGAGCAG GTATACTGGT CGTGTGCCCGA
 15481 CATGATGCAA GACCCCGTGA CCTTCCGCTC CACGCGCCAG ATCAGCAACT TTCCTGGTGT
 15541 GGGCGCGCAG CTGTTGCCCG TGCACTCCAA GAGCTTCTAC AACGACCAGG CCGCTCTACT
 15601 CCAGCTCATC CGCCAGTTTA CCTCTCTGAC CCACGTGTTT AATCGCTTTC CCGGAACCA
 15661 GATTTTGGCG CGCCCGGCCG CCCCACCAT CACCACCGTC AGTGAAGAGC TTCTCGCTCT
 15721 CACAGATCAC GGGACGCTAC GCCTGCGCAA CAGCATCGGA GGAGTCCAGG GAGTGACCAT
 15781 TACTGACGCC AGACGCGGCA CCTGCCCTTA CGTTTACAAG GCCCTGGGCA TAGTCTCGCC
 15841 GCGCGTCTTA TCGAGCCGCA CTTTTGAAG AAGCATGTCC ATCCTTATAT CGCCACGAA
 15901 TAACACAGCG TGGGGCTGCG GCTTCCCAA GCAAGATGTT GCGCGGGGCA AGAAGCGCTC
 15961 CGACCAACAC CCAGTGCGCG TGCGCGGGCA CTACCGCGCG CCTTGGGCGG CGCACAAAGC
 16021 CGCGCCACT GGGCGCACCA CCGTCGATGA CGCCATCGAG CGCGTGTGTG AGGAGGCGCG
 16081 CAATACACG CCCACGCCGC CGCCAGTGTG CACCGTGGAG GCGGCCATTG AGACCTGTGT
 16141 CGCGCGAGGCC CGCGCTACG CTAATAAGAA GAGACGCGCG AGGCGCGTAG CACGCTGCCA
 16201 CGCGCGCGA CCGCGCACTG CCGCCCAACG CGCGCGCGCG GCCCTGCTTA ACCGCGCAG
 16261 TCGCACCGCG CGACGGGCGG CCATGCGAGC CGCTCGAAGG CTGGCCCGCG GTATTGTCTC
 16321 TGTGCCCCCC AGGTCACGCG GACGAGCGCG CGCCGACGCA GCGCGCGGCA TTATGTGCTAT
 16381 GACTCAGGGT CGCAGGGGCA ACGTGTA CTG GGTGCGCGAC TCGGTTAGCG GCCTGCGCGT
 16441 GCCCGTGC CGCCGCCCCC CGCGCAACTA GATTGCAATA AAAA CTA CTA TAGACTCGTA
 16501 CTGTTGTATG TATCCAGCGG CGCGCGCGCG CATCGAAGCT ATGTCCAAAG GCAAAATCAA
 16561 AGAAGAGATG CTCAGGTCA TCGCGCGGGA GATCTATGGC CCCCAGAAAG AGGAAGAGCA
 16621 GGATTACAAG CCCCAGAAAG TAAAGCGGGT CAAAAGAAAG AAGAAAGATG ATGATGATGA
 16681 TGAACCTGAC GACGAGGTGG AACTGTTGCA CGCGACCGCG CCCAGCGCAC GGGTACAGTG
 16741 GAAAGGTGCA CGCGTAAGAC GTGTTTGGG ACCCGGCAAC ACCGTATGCT TTACGCCCGG
 16801 TGAGCGCTTC ACCCGCACTT ACAAGCGCGT GTATGATGAG GTGTACGCGC ACGAGGACCT
 16861 GTCTGAGCAG GCGCAACGAG GCCTCGGGGA GTTTGCTTAC GGAAGCGCGC ATAGGACAT
 16921 GCTGGCGTTG CCGCTGGACG AGGGCAACCC AACACCTAGC CTAAGCGCCG TGACACTGCA
 16981 CGAGGTGCTG CCCGCGCTTG CACCGTCCGA AGAAAAGCGG GGCCATAAGC GCGAGTCTGG
 17041 TGACTTGGCA CCGACCGTGC AGCTGATGTT AGCCCAAGCGT CAGCGACTGG AAGATGTCTT
 17101 GGAATAAATG ACCGTGGAGC CTGGGTGGA GCCCGAGGTC CGCGTGC CGC CAATCAAGCA
 17161 GGTGGCACCG GGAAGTGGCG TGACAGACCT GGACGTTTCA ATACCCACCA CCAAGTACAC
 17221 TAGTATTGCT ACTGCCACAG AGGGCATGGA GACACAAAGC TCCCGCGTGT CCTCGGCGGT
 17281 GGCAGATGCC GCGGTGACAG CGGCCGCTGC GGCCGCGTCC AAGACCTCTA CGSAGGTGCA
 17341 ACGGACCCCG TGAGTGTTCG GTGTTTCAAG CCCC CGCGCT CCGCGCGCTT CAGGAAGATA
 17401 CGCGCGCGCG AGCGCGCTAC TGCCCGAATA TGCCCTACAT CTTTCCATGT CGCTCGGCGT
 17461 CGGCTATCGT GGCTACACCT ACCGCCCGAG AAGACGAGCA ACTACCCGAC GCCGAACCC
 17521 CACTGGAACC CGCGCGCGCG GTGCCGCTCG CAGCGCCGTC CTGGCCCGCA TTTTCGTGCG
 17581 CAGGGTGGCT CGCGAAGGAG CGAGGACCTT GGTGTGCTCA ACGCGCGCTT ACCACCCGAG

FIG. 7G

49/92

17641 CATCGTTTAA AAGCCGGTCT TTGTGGTTCT TGCGATATG GCCCTCACCT GCCGCTCCG
 17701 TTTCOCGGTG CCGGATATCC GAGGAAGAAT GCACCGTAGG AGGGGCATGG CCGGCCACGG
 17761 CCTGACGGGC GGCATGCGTC GTGCGCACCA CCGCGGCGG CGCGCGTCGC ACCGTCGCAT
 17821 CGCGCGCGGT ATCCTGCCCC TCCTTATATCC ACTGATCGCC GCGGCGATGG CCGCGGTGCC
 17881 CGGAATTGCA TCCGTGGCCT TGCAGGCGCA GAGACACTGA TTAATAACAA GTTACATGTG
 17941 GAAAAATCAA AATAAAAGTC TGGACTCTCA CGCTCGCTTG GTCTGTGAAC TATTTGTAG
 18001 AATGGAAGAC ATCAACTTTG CGTCACTGGC CCCGCGACAC GGCTCGGCGC CGTTCATGGG
 18061 AAATCGGCAA GATATCGGCA CCAGCAATAT GAGCGGTGGC GCCTTCAGCT GGGGCTCGCT
 18121 GTTGCGCGGC ATTAAAAATT TCGTTCGCC CGTTAAGAAC TATGGCAGCA AAGGCTGGAA
 18181 CAGCAGCACA GGCCAGATGC TGAGGGACAA GTTGAAGAG CAAATTTCC AACAAAAGGT
 18241 GGTAGATGGC CTGGCCTCTG GCATTAGCGG GGTGGTGGAC CTGGCCAAAC AGGCAGTGCA
 18301 AAATAAGATT AACAGTAAGC TTGATCCCGC CCCTCCCGTA GAGGAGCCTC CACCGCGCTT
 18361 GGAGACAGTG TCTCCAGAGG GCGGTGGCGA AAAGCGTCCG CGACCCGACA GGGAGAAGAC
 18421 TCTGGTGACG CAAATAGACG AGCCTCCCTC GTACAGGAGG GCATTAAGCG AAGGCTGGCC
 18481 CACCACCCGT CCCATCGCGC CCATGGCTAC CGGAGTGCTG GGCCAGTACA CACCCGTAAAC
 18541 GCTGGACCTG CCTCCCCCGC CCGACACCCA GCAGAAACCT GTGCTGCCAG GCCCGTCCGC
 18601 CGTTGTGTGA ACCCGTCCTA GCGCGCGCTC CCTGCGCCGC GCGCGCACGG CTGCGCGATC
 18661 GTTGCGCGGC GTAGCCAGTG GCAACTGGCA AAGCACACTG AACAGCATCG TGGGTTTGGG
 18721 GGTGCAATCC CTGAAGCGCC GACGATGCTT CTGATAGCTA ACGTGTGCTA TGTGTGTGAT
 18781 GTATGCGTGC ATGTGCGCGC CAGAGGAGCT GCTGAGCGCG CGCGCGCCCG CTTTCGAAGA
 18841 TGGCTACCCC TTCGATGATG CCGCATGTGC CTTCATGCTA CATCTCGGCG CAGGACCGCT
 18901 CGGAGTACCT GAGCCCCGGG CTGGTGCACT TCGCCCCGCG CACCGAGACG TACTTCAGCC
 18961 TGAATAACAA GTTTAGAAAC CCCACGGTGG CGCCTACGCA CGACGTGACC ACAGACCGGT
 19021 CTCAGCGGTT GACGCTGCGG TTCAATCCCG TGGAACCGGA GGATACTGCG TACTCGTATA
 19081 AGGCGCGGTT CACCTTAGCT GTGGGTGATA ACCGTGTGCT AGACATGGCT TCCACGTACT
 19141 TTGACATCCG CGCGGTGCTG GACAGGGGCC CTACTTTTAA GCCCTACTCT GGCACGTGCT
 19201 ACAACGCACT GGGCCCCAAG GGTGCCCCCA ACTCGTGGCA GTGGGAACAA AATGAACATG
 19261 CAAAGTGGTA TGCTCAAGAA CTTGACGAAG AGGAGAATGA AGCCAAATGA CTCTAGGCGC
 19321 GAGAACAGGA ACAAGCTAAG AAAACCCATG TATATGCCCA GGCTCCACTG TCCGGAATAA
 19381 AAATAACTAA AGAAGGTCTA CAAATAGGAA CTGCGGACGC CACAGTAGCA GGTGCCGGCA
 19441 AAGAAATTTT CGCAGACAAA ACTTTTCAAC CTGAACCACA AGTAGGAGAA TCTCAATGGA
 19501 ACCGAGCGGA TGCCACAGCA GCTGGTGAA GGGTTCTTAA AAGACAACTC CCCATGAAC
 19561 CCTGCTATGG CTCATACGCT AGACCCACCA ATTCCAACGG CGGACAGGGC GTTATGTTG
 19621 AACAAATTTG TAAATTGGAA AGTCAAGTCG AAATGCAATT TTTTCCACA TCCACAAATG
 19681 CCACAAATGA AGTTAACAT ATACAACCAA CAGTTGTATT GTACAGGAA GATGTAAACA
 19741 TGGAATCTCC AGATACTCAT CTTTCTTATA AACCTAAAAA GGGGGATAAA AATGCCAAAG
 19801 TCAATGCTTG ACACAAGCA ATGCCAAACA GACCAAAATTA CATTTGCTTT AGAGACAATT
 19861 TTAATGTGCT CATGTATTAC AACAGCACAG GTAACATGGG TGTCTGTGCT GGTGAGCAT
 19921 CGCAGTTGAA CGCTGTTGTA GATTGTGCAAG ACAGAAACAC AGAGCTGTCC TACCAGCTTT
 19981 TGCTGTATTC AATTGGCGAC AGAACAGAT ACTTTTCAAT GTGGAATCAA GCTGTTGACA
 20041 GCTATGATCC AGATGTCAGA ATTATTGAGA ACCATGGAAC TGAGGATGAG TTGCCAAATT
 20101 ATTGCTTTCC TCTTGGTGA ATTGGGATTA CTGACACTTT TCAAGCTGTT AAAACAACCTG

FIG. 7H

50/92

20161 CTGCTAACGG GGACCAAGGC AATACTACCT GGCAAAAAGA TTCAACATTT GCAGAAGCGA
 20221 ATGGAATAGG GGTGGGAAT AACTTTGCCA TGGAAATAA OCTGAAATGCC AACCTATGGA
 20281 GAAATTTCTT TTACTCCAAT ATTGGCGTGT ACCTGCCAGA CAAGCTAAAA TACAACCCCA
 20341 CCAATGTGGA AATATCTGAC AACCCCAACA CCTACGACTA CATGAACAAG CGAGTGGTGG
 20401 CTCTGGGGCT TGTAGACTGC TACATTAAC TTGGGGCGCG CTGGTCTCTG GACTACATGG
 20461 ACAACGTTAA TCCCTTTAAC CACCACCGCA ATGGGGCGCT GCGTTACCGC TCCATGTTGT
 20521 TGGGAAACGG CCGCTACGTG CCCTTTTACA TTACAGTGGC CCAAAAGTTT TTTGCCATTA
 20581 AAAACCTCCT CCTCCTGCCA GGCTCATACA CATATGAATG GAACCTCAGG AAGGATGTTA
 20641 ACATGGTTCT GCAGAGCTCT CTGGGAAACG ACCTTAGAGT TGACGGGGCT AGCATTAAAT
 20701 TTGACAGCAT TTGCTTTTAC GCCACCTTCT TCCCATGGC CCACAACAGC GCCTCCACGC
 20761 TGGAGGCCAT GCTCAGAAAT GACACCAACG ACCAGTCTTT TAATGACTAC CTTTCGCGCG
 20821 CCAACATGCT ATATCCCATC CCGGCCAACG CCACCAACGT GCCCATCTCC ATCCCATTCG
 20881 GCAACTGGGC AGCATTTCGC GGTTCGGGCT TCACACGCTT GAAGACAAAG GAAACCCCTT
 20941 CCCTGGGATC AGGCTACGAC CCTTACTACA CCTACTCTGG CTCCATACCA TACCTTGAGG
 21001 GAACCTCTTA TCTTAATCAC ACCTTTAAGA AGGTGGGCAT TACTTTTACG TCTTCTGTTA
 21061 GCTGGCCGGG CAACGACCGC CTGCTTACTC CCAATGAGTT TGAGATTAGC GCCTCAGTTG
 21121 ACGGGGAGGG CTATAACGTA GCTCAGTGCA ACATGACAAA GGACTGGTTC CTAGTGCAGA
 21181 TGTGTGGCCAA CTCACAATAT GGCTACCAGG GCTTCTACAT TCCAGAAAGC TACAAGAAC
 21241 GCATGTACTC GTTCTTCAGA AACTTCCAGC CCATGAGCCG GCAAGTGGTG GACGATACTA
 21301 AATACAAAGA TTATCAGCAG GTTGGAAATTA TCCACCAGCA TAACAACCTCA GGCTTCGTAG
 21361 GCTACTCTGC TCCACCATG CGCGAGGGAC AAGCTTACCC CGCTAATGTT CCTTACCCAC
 21421 TAATAGGCCAA AACCGCGGTT GATAGTATTA CCCAGAAAAA GTTTCCTTGC GACCGACACC
 21481 TGTGGCGCAT CCCCTTCTCC AGTAACCTTA TGTCCATGGG TGGCTTCACA GACCTGGGCC
 21541 AAAACCTTCT CTACGCAAC TCCGCCACG CGCTAGACAT GACCTTTGAG GTGGATCCCA
 21601 TGGACGAGCC CACCCTTCTT TATGTTTGTG TTGAAGTCTT TGACGTGGTC CGTGTGCACC
 21661 AGCCGCAACG CGGCGTCATC GAGACCGTGT ACCTGCGCAC GCCTTCTCGC GCCGCAACG
 21721 CCACAACATA AAGAAGCAAG CAACATCAAC AACAGCTGCC GCCATGGGCT CCAGTGAGCA
 21781 GGAAGTGAAA GCCATTGTCA AAGATCTTGG TTGTGGGCCA TATTTTFTGG GCACCTATGA
 21841 CAAGCGCTTC CAGGCTTTG TTTCCCAACA CAAGCTCGCC TGGCCCATAG TTAACACGGC
 21901 CGGTCCGAGG ACTGGGGGCG TACACTGGAT GGCCTTTGCC TGGAAACCGG GCTCAAAAA
 21961 ATGCTACCTC TTTGAGCCCT TTGGCTTTTC TGACCAACGT CTCAGCAGG TTTACCAAGT
 22021 TGAGTACGAG TCACTCTGCG GCGGTAGGCG CATTCCTCT TCCCCGACC GCTGTATAAC
 22081 GCTGGAAGAAG TCCACCCAAA CGGTGAGGGG GCCCAACTCG GCCGCTGTG GCTATTCTG
 22141 CTGCATGTTT CTCCACGCTT TTGCCAACTG GCCCAAACT CCCATGGATC ACAACCCAC
 22201 CATGAACCTT ATTACCGGGG TAOCCAACTC CATGCTTAAC AGTCCCCGAC TACAGCCAC
 22261 CCTGCGCGCG AACCAAGAAC AGCTCTACAG AGCTCTACAG CGCCACTCGC CTTCTTCG
 22321 CAGCCAAGT GCGCAAAATTA GGAGCGCCAC TTCTTTTGTG CACTTGAATA ACATGTAAAA
 22381 ATATGTACT AGGAGACACT TTCAATAAAG GCAAAATGTT TATTTTGTAC ACTCTCGGTT
 22441 GATTATTTAC CCCACCCCTT GCGGCTGCGC CCGTTTAAAA ATCAAGGGGG TTTGCGCGCG
 22501 CATCGCTATG CGCCACTGGC AGGGACACGT TCGCATACTG GTGTTTAGTG CTCCACTTAA
 22561 ACTCAGGCAC AACCATCCGC GGCAGCTCGG TGAAGTTTTC ACTCCACAGG CTGCGCACCA
 22621 TCACCAACGC GTTTAGCAGG TCGGGCGCGC ATATCTTGAA GTGCGAGTTG GGGCTCCGC

FIG. 71

51/92

22681 CCTGCGCGCG CGAGTTGCGA TACACAGGGT TACAGCACTG GAACACTATC AGCGCCGGGT
 22741 GGTGCACGCT GGCAGCAGC CTCTTGTTCG AGATCAGATC CGCGTCCAGG TCCTCCGCGT
 22801 TGCTCAGGGC GAACGGAGTC AACTTTGGTA GCTGCCTTCC CAAAAAGGGT GCATGCCCGAG
 22861 GCTTTGAGTT GCACTGCGAC CGTAGTGGCA TCAGAAGGTG ACCGTGCCCA GTCTGGCGGT
 22921 TAGGATACAG CGCCTGCATG AAAGCCTTGA TCTGCTTAAA AGCCACCTGA GCCTTTGCGC
 22981 CTTCAGAGAA GAACATGCCG CAAGACTTGC CGGAAAACGT ATTGGCCGGA CAGGCCCGGT
 23041 CATGCACGCA GCACCTTGGC TCGGTGTGG AGATCTGCAC CACATTTCCG CCCACCCGGT
 23101 TCTTCACGAT CTTGGCCTTG CTAGACTGCT CCTTCAGCGC GCGCTGCCCG TTTTTCGCTG
 23161 TCACATCCAT TTCAATCACG TGTCTTTAT TTATCATAAAT GCTCCCGTGT AGACACTTAA
 23221 GCTCGCCTTC GATCTCAGCG CAGCGGTGCA GCCACAACGC GCAGCCCGTG GGCCTGFTGT
 23281 GCTTTGAGGT TACCTCTGCA AACGACTGCA GGTACGCCCTG CAGGAATCGC CCCATCATCG
 23341 TCACAAAGGT CTGTGTGCTG GTGAAGGTCA GCTGCAACCC GCGGTGCTCC TCGTTTAGCC
 23401 AGGTCTTGCA TACGCCCGCC AGAGCTTCCA CTGTGTGTCAG CAGTAGCTTG AAGTTTGCTT
 23461 TTAGATCGTT ATCCACGTGG TACTTGTCCA TCAACGCGCG CGCAGCCTCC ATGCCCTTCT
 23521 CCAACGCA GAAGATCGCG AGGCTCAGCG GGTTTATCAC CGTGCTTTCA CTTTCCCGCT
 23581 CACTGGACTC TTCCTTTTCC TCTTGCAATCC GCATACCCCG CGCCACTGGG TCGCTTTCAT
 23641 TGACCGCGCG CACCGTGC GC TTAACCTCCCT TGCCTGCTTT GATTAGCAAC GGTGGGTTCG
 23701 TGAAACCCAC CATTTGTAGC GCCACATCTT CTCTTTCTTC CTCGCTGTCC ACGATCACCT
 23761 CTGGGATGG CGGGCGCTCG GGCTTGGGAG AGGGGCGCTT CTTTTCCTTT TTTGACGCAA
 23821 TGCCCAAACT CGCGCTGAG GTGATGSGCC GCGGGCTGGG TGTGCGCGGC ACCAGCGCAT
 23881 CTTGTGACGA GTCTTCTTCG TCCTCGGACT CGAGACGCGC CCTCAGCGCG TTTTTCGGGG
 23941 GCGCGCGGGG AGCGCGCGGC GACGCGACG GGGACGAGAC GTCTCCCATG GTTGTGGGAC
 24001 GTCGCGCGCG ACCGCGTCCG CGCTCGGGGG TGGTTTCGCG CTGCTCCCTC TCCCGAGCTG
 24061 CCATTTCTCT CTCCTATAGG CAGAAAAAGA TCATGGAATC AGTCGAGAAG GAGGACAGCC
 24121 TAAACGCCCC CTTTGAAGTC GCCACCAACG CCTCCACCGA TGCCGCCAAC GCGCTACCA
 24181 CCTTCCCGCT CGAGGCACCC CCGCTTGAGG AGGAGGAAGT GATTATCGAG CAGGACCCAG
 24241 GTTTTGTAA GCAAGACGAC GAAGATCGCT CAGTACCAAC AGAGGATAAA AAGCAAGACC
 24301 AGGACGACGC AGAGGCAAA C GAGGAAACAG TGGGGCGGGG GGACCAAAAG CATGGCGACT
 24361 ACTTAGATGT GGGAGACGAC GTGCTGTGA AGCATCTGCA GCGCGAGTGC GCCATTATCT
 24421 GCGACGCGTT GCAAGAGCGC AGCGATGTGC CCCTCGCCAT AGCGGATGTC AGCCTTGCTT
 24481 ACGAAGCGCA CCTGTCTCA CCGCGCGTAC CCCCCAACG CCAAGAAAAC GGCACATGCG
 24541 AGCCCAACCC GCGCCTCAAC TTCTACCCCG TATTTCGCGT GCCAGAGGTG CTTGCAACCT
 24601 ATCACAATCT TTTCCAAAAC TGCAAGATAC CCCATATCCT CCGTGCCAAC CGCAGCCGAG
 24661 CGGACAGCA GCTGGCCTTG CGGCAAGGCG CTGTATACCC TGATATCGCC TCGCTCGAGC
 24721 AAGTGCCAAA AATCTTTGAG GGTCTTGAGC GCGACGAGAA GCGCGCGGCA AAGCCTCTGC
 24781 AACAAAGAAA CAGCGAAAT GAAAGTCACT GTGGAGTGCT GGTGGAACCT GAGGCTGACA
 24841 ACGCGCGCCT AGCGGTGCTG AAACGAGCA TCGAGGTCAC CCACCTTGGC TACCCTCGAGC
 24901 TTAACCTACC CCCCAAGGTT ATGAGCAGC TCATGAGCGA GCTGATCGTG GCGCGTGCAC
 24961 GACCCCTGGA GAGGGATGCA AACTTGCAAG AACAAACCGA GGAGGGCCTA CCGCGAGTTG
 25021 CGGATGAGCA GCTGGCGCGC TGGCTTGAGA CGCGCGAGCT TGCCGACTTG GAGGAGCGAC
 25081 GCAAGCTAAT GATGGCCGCA GTGCTTGTGA CCGTGGAGCT TGAGTGCATG CAGCGTTCCT
 25141 TTGCTGACCC GGAGATGCAG CGCAAGCTAG AGGAACGTT GCATACACCC TTTGCGCAGG

FIG. 7J

52/92

25201 GCTACGTGCG CCAGGCCTGC AAAATTTCCA ACCTGGAGCT CTGCAACCTG GTCTCTTACC
 25261 TTGGAAATTTT GCACGAAAC CGCCTTGGGC AAAACGTGCT TCAATCCAGC CTCGAAGGGG
 25321 AGGCGCGCGC GCACTACGTC CGGACTGCG TTTACTTATT TCTGTGCTAC ACCTGGCATA
 25381 CGGCCATGGG CGTGTGGCAG CAGTGCTGG AGGAGCGCAA CCTGAAGGAG CTGCAGAGAC
 25441 TGCTAAAGCA AACTTTGAAG GACCTATGGA CGCCTTCAA CGAGCGCTCC GTGGCGCGCG
 25501 ACCTGGCGGA CATTATCTTC CCCGAACGCC TGCTTAAAC CCTGCAACAG GGTCTGCCAG
 25561 ACTTCACCAG TCAAGCATG TTGCAAACT TTAGGAACCT TATCCTAGAG CGTTCAGGAA
 25621 TTCTGCCCGC CACCTGCTGT GCGCTTCCTA GCGACTTTGT GCCCATTAAG TACCGTGAAT
 25681 GCCCTCGGCC GCTTTGGGGT CACTGCTACC TTCTGAGCT AGCCAACTAC CTTGCCCTACC
 25741 ACTCGACAT CATGGAAGAC GTGAGCGGTG ACGCCCTACT GGAAGTGTAC TGTGCTGTGA
 25801 ACCTATGCAC CCCGCACCGC TCCTTGGTCT CGCAATTCACA ACTGCTTAGC GAAAGTCAAA
 25861 TTATCGGTAC CTTTGAGCTG CAGGGTCCCT GCCTGACGA AAGTCCGCG GCTCCGGGGT
 25921 TGAACCTCAC TCCGGGCGTG TGGACGTGCG CTTACCTTCG CAAATTTGTA CCTGAGGACT
 25981 ACCACGCCCA CGAGATTAGG TTCTACGAAG ACCAATCCCG CCCGCCAAAT CGCGAGCTTA
 26041 CGCCCTCGCT CATTACCAG GGCACATCC TTGGCCAAAT GCAAGCATT AACAAAGCC
 26101 GCCAAGATT TCTGTACGA AAGGGACGGG GGGTTACTT GGACCCCGAG TCCGGCGAGG
 26161 AGCTCAACCC AATCCCCCG CGCGCGCAG CCTATCAGCA CGCCCGGGCC CTTGCTTCCC
 26221 AGGATGGCAG CCAAAAGAA GCTGCAGCTG CGCGCGCCG CACCCACGGA CGAGGAGGAA
 26281 TACTGGGACA GTACGSCAGA GGAGTTTGT GACGAGGAGG AGGAGATGAT GGAAGACTGG
 26341 GACAGCCTAG ACGAGGAAGC TTCCGAGGCC GAAGAGGTGT CAGACGAAC ACCGTCACCC
 26401 TCGGTCGCAT TCCCTCGCC GCGCGCCAG AAATCGGCAA CCGTTCCAG CATTGCTACA
 26461 ACTTCCGCTC CTCAGCGGCC GCGGCACCTG CCGTTCGCC GACCCAACCG TAGATGGGAC
 26521 ACCACTGGAA CCAGGGCGCG TAAGTCTAAG CAGCCGCGCG CGTTAGCCCA AGAGCAACAA
 26581 CAGCGCCAAG GCTACCGCTC GTGGCGCGTG CACAAGAACG CCATAGTTGC TTGCTTGCAA
 26641 GACTGTGGGG GCAACATCTC CTTGCGCCCG CGCTTTCTTC TCTACCATCA CGCGGTGGCC
 26701 TTCCCCGTA ACATCTGCA TTACTACCGT CATCTCTACA GCCCCTACTG CACCGGCGCG
 26761 AGCGGCAGCA ACAGCAGCG CCACGCAGAA GCAAAGGCGA CCGGATAGCA AGACTCTGAC
 26821 AAAGCCCAAG AAATCCACAG CGCGGCGAGC AGCAGGAGGA GGAGCACTGC GTCTGGCGCC
 26881 CAACGAACCC GTATCGACC GCGAGCTTAG AAACAGGATT TTTCCTACT TGTATGCTAT
 26941 ATTCTACAG AGCAGGGGCG AAGAACAAGA CTGAAATAA AAAAACAGGT CTTGCGCTC
 27001 CCTCACCGC AGCTGCTGT ATCACAAGA CGAAGATCAG CTTGGGCGCA CGCTGGAGA
 27061 CGCGGAGGCT CTCTTCAGCA AATACTGCGC GCTGACTTT AAGGACTAGT TTGCGGCCCT
 27121 TTCTCAAAAT TAAGCGCGAA AACTACGTCA CTCCAGCGG CCACACCGCG CGCAGCACCC
 27181 TGTCGTGAGC GCCATTATGA GCAAGGAAAT TCCACGCCC TACATGTGGA GTTACCAGCC
 27241 ACAAAATGGA CTTGCGGCTG GAGCTGCCCA AGACTACTCA ACCCGAATAA ACTACATGAG
 27301 CGCGGAGACC CACATGATAT CCGGGGTCAA CGGAATCCCG GCCACCGAA ACCGAATTCT
 27361 CCTGAAACAG CGGCTATTA CCACACACCC TCGTAATAAC CTTAATCCCC GTAGTTGGCC
 27421 CGTGCCTCG GTGTACCAGG AAAGTCCCGC TCCACCACT GTGCTACTTC CCAGAGACCG
 27481 CCAGGCCGAA GTTCAGATGA CTAACCTCAG GCGCAGCTT CGGGGCGGCT TTCTGCACAG
 27541 GGTGCGGTG CCGGCGCAG GTATACTCA CTGAAATCT AGAGGGCGAG GTATTTCAGCT
 27601 CAACGACGAG TCGGTGAGCT CCTCTCTGG TCTCGTCCG GACGGGACAT TTCAGATCGG
 27661 CGGCGCTGGC CGCTCTTCAT TTACGCCCGC TCAGGCGATC CTAACCTCTG AGACCTCTGC

FIG. 7K

53/92

27721 CTCGGAGCCG CGCTCCGGAG GCATTGGAAC TCTACAATTT ATTGAGGAGT TCGTGCCCTTC
 27781 GGTTTACTTTC AACCCCTTTT CTGGACCTCC CGGCCACTAC CCGGACCAGT TTAATCCCAA
 27841 CTTTGACGCG GTAAAGACT CGGCGGACGG CTACGACTGA ATGACCAGTG GAGAGGCAGA
 27901 GCAACTGCGC CTGACACACC TCGACCACTG CCGCCGCCAC AAGTGCTTTG CCGCGGCGTC
 27961 CGGTGAGTTT TGTTACTTTG AATTGCCCGA AGAGCATATC GAGGGCCCGC CGCAGCGCGT
 28021 CGGCTCACC ACCCAGGTAG AGCTTACACG TAGCCTGATT CGGGAGTTTA CCAAGCGCCC
 28081 CCGTCTAGTG GAGCGGGAGC GGGGTCCCTG TGTCTGACC GTGGTTGCA ACTGTCTTAA
 28141 CCCTGGATTA CATCAAGATC TTTGTTGTCA TCTCTGTGCT GAGTATAATA AATACAGAAA
 28201 TTAGAATCTA CTGGGGCTCC TGTCGCCATC CTGTGAACGC CACCGTTTTT ACCCACCCAA
 28261 AGCAGACCAA AGCAAACCTC ACCTCCGGTT TGCACAAGCG GGCCAATAAG TACCTTACCT
 28321 GGTACTTTAA CGGCTCTTCA TTTGTAATTT ACAACAGTTT CCAGCGAGAC GAAGTAAGTT
 28381 TGGCACACAA CCTTCTCGGC TTCAACTACA CCGTCAAGAA AACACCACC ACCACCTCC
 28441 TCACCTGCGC GGAACGACG AGTGCCTCAC CGGTTGCTGC GCCCACCTT ACAGCCTGAG
 28501 CGTAACACAGA CATTACTCCC ATTTTCCCAA AACAGGAGGT GAGCTCAACT CCCGGAACCT
 28561 AGGTCAAAAA AGCAATTTTG GGGGTGCTGG GATTTTTAA TTAAGTATAT GAGCAATTTA
 28621 AGTAACCTTA CAAGCTTGTC TAATTTTCTT GGAATTTGGG TCGGGGTTAT CCTTACTCTT
 28681 GTAATTCGTG TTATTCCTAT ACTAGCACT CTGTGCCCTA GGGTTGCCGC CTGCTGCAGC
 28741 CACGTTTGTG CCTATTGTCA GCTTTTAAA CGCTGGGGGC GACATCCAAAG ATGAGCTACA
 28801 TGATTTTAGG CTTGCTCGCC CTTCGCGCAG TCTGCAGCGC TGCCAAAAAG GTTGAGTTTA
 28861 AGGAACACAG CTGCAATGTT ACATTTAAAT CAGAAGCTAA TGAATGCAC TCTCTTATAA
 28921 AATGCACCA AGAACAATGA AAGCTTATTA TTGCCACAA AGACAAAATT GGCAGATATG
 28981 CTGTATATGC TATTGGCAG CCAGGTGACA CTAACGACTA TAATGTCACA GTCTTCCAAG
 29041 GTGAAAATCG TAAACTTTTT ATGTATAAAT TTCCATTTTA TGAATGTGC GATATTACCA
 29101 TGTACATGAG CAAACAGTAC AAGTTGTGGG CCCACRAAA GTGTTTAGAG AACCATGGCA
 29161 CCTTTTGTTC CACCGCTCTG CTTATTACAG CGCTTGCTTT GGTATGTACC TTACTTTTATC
 29221 TCAAAATACA AAGCAGACGC AGTTTATTTG ATGAAAAGAA AATGCCCTTA TTTTCCGCTT
 29281 CCTTGTATCT CCTTGGACAA TTTACTCTAT TGGGGATATG CGCCAGCGCG GAAAGATTAT
 29341 ACCCAACAAC TTCAAATCAA ACTTTCTTGG ACGTTAGCGC CTGACTTCTG CCAGCGCCTG
 29401 CACTGCAAA TTTGATCAAA CCAGCTTACG CTTGCTGTCT CCAGAGATGA CCGGCTCAAC
 29461 CATCGCGCCC ACAACGGACT ATCGCAACAC CACTGCTTAC GGACTAAAAT TGCCCTTAAA
 29521 TTTACCCCAA GTTCATGCTT TTGTCAATGA CTGGGCGAGC TTGGGCTATG GGTGTTTTC
 29581 CATAGCGCTT ATGTTTGTTT GCCTTATTAT TATGTGCTT ATTTGTTGCG TAAAGCGCAG
 29641 ACGCGCCAGA CCCCCATCT ATAGGCCCTAT CATTTGCTCT AACCCACACA ATGAAAAAAT
 29701 TCATAGATTG GACGGTCTCA AACCATTGTC TCTTCTTTTA CAGTATGATT AAATGAGACA
 29761 TGATTCCTCG AGTCCTTATA TTATTGACCC TTGTTGCGCT TTTCTGTGCG TGCTCTACAT
 29821 TGGCTGCGGT CGCTCACATC GAAGTAGATT GCATCCACCC TTTCACAGTT TACCTGCTTT
 29881 ACGGAATTTG CACOCCTATC CTCATCTGCA GCTCGCTCAC TGTAGTCACT GCCTTCATTT
 29941 AGTTCATGTA CTGGATTTGT GTGCGCATTG CGTACCTTAG GCACCAATCG CAATACAGAG
 30001 ACAGGACTAT AGCTGATCTT CTCAGAATTC TTTAATPATG AAACGGATTG TCACTTTTGT
 30061 TTTGCTGATT TTCTGCGCCC TACCTGTGCT TTGCTGCCAA ACCTCAGCGC CTCCCAAAG
 30121 ACATATTTCC TGCAGATTCA CTCAAAATG GAACATTTCC AGCTGCTACA ACAAACAGAG
 30181 CGATTTGTCA GAAGCTTGGT TATACGCCAT CATCTCTGTC ATGGTTTTTT CGAGTACCAT

FIG. 7L

54/92

30241 TTTTGCCCTA GCCATATACC CATACCTTGA CATTGGTTGG AATGOCATAG ATGCCATGAA
 30301 CCACCCTACT TTCCAGCGC CCATGTTCAT ACCACTGCAA CAGGTTATTG CCCCAATCAA
 30361 TCAGCCTCGC CCCCTTCTCT CCACCCCCAC TGAGATTAGC TACTTTAATT TGACAGGTGG
 30421 AGATGACTGA ATCTCTAGAT CTAGAATTGG ATGGAATTAAC CACCAGACAG CGCCTACTAG
 30481 AAAGGCGCAA GCGGCGTCC GAGCGAGAAC GCCTAAACA AGAAGTTGAA GACATGGTTA
 30541 ACCTGCACCA GTGTAAAAGA GGTATCTTTT GTGTGGTCAA GCAGGCCAAA CTTACCTACG
 30601 AAAAACCAC TACCGGCAAC CGCCTTAGCT ACAAGCTACC CACCAGCGC CAAAACTGG
 30661 TGCTTATGGT GGGAGAAAAA CCTATCACC TCACCCAGCA CTCGGCAGAA ACAGAGGGT
 30721 GCCTGCACCT CCCCTATCAG GGTCCAGAGG ACCTCTGCAC TCTTATTAAA ACCATGTGTG
 30781 GCATPAGAGA TCTTATTCCA TTCAACTAAC AATAAACACA CAATAAATTA CTTACTTAAA
 30841 ATCAGTCAGC AAATCTTTGT CCAGCTTATT CAGCATCACC TCCTTTCCCT CCTCCCACT
 30901 CTGGTATTTC AGCAGCCTTT TAGCTGCGAA CTTTCTCCAA AGTCTAAATG GGATGTCAA
 30961 TTCCTCATGT TCTTGTCCCT CCGACCCAC TATCTTCATA TTGTTGAGA TGAACGCGC
 31021 CAGACGCTCT GAAGACACCT TCAACCTGTG GTACCCATAT GACACGGAAA CCGGCCCTCC
 31081 AACTTGTGCT TTCTCTTACC CTCCCTTTGT GTCGCCAAAT GGGTTCCAAG AAGTCCCCC
 31141 CGAGTGTCTT TCTTTCGCTC TTTCAGAAC TTTGGTTACC TCACACGGCA TGTCTGGCT
 31201 AAAAATGGGC AGCGGCCCTGT CCCCTGGATCA GGCAGGCAAC CTTACATCAA ATACAATCAC
 31261 TGTTTCTCAA CCGCTAAAAA AAACAAGATC CAATATAACT TTGGAACACT CCGGCCCTCT
 31321 TACAGTCAGC TCAGGCGCCC TAACCATGCG CACAACCTCG CTTTGTGTGG TCTCTGACAA
 31381 CACTCTTACC ATGCAATCAC AAGCACCGCT AACCGTGCAA GACTCAAAAC TTAGCATTCG
 31441 TACCAAAGAG CCACCTTACAG TGTTAGATGG AAAAAGTGCC CTGCAGACAT CAGCCCCCCT
 31501 CTCCTGCACT GATAACAACG CCCTCACTAT CACTGCCTCA CCTCTCTTAT CTACTGCAAA
 31561 TGGTAGTCTG GCTGTTACCA TGGAAAACCC ACTTTACAAC AACAAATGGAA AACTTGGGCT
 31621 CAAAATTTGG GGTCCCTTTC AAGTGGCCAC CGACTCACAT GCACTAACAC TAGGTACTGG
 31681 TCAGGGGGTT GCAGTTCATA ACAATTTGCT ACATACAAA GTTACAGGCG CAATAGGGTT
 31741 TGATACATCT GGCAACATGG AACTTAAAC TGGAGATGGC CTCTATGTGG ATAGCGCCGG
 31801 TCTTAACCAA AAACCTACATA TTAATCTAAA TACCACAAA GGCTTTGCTT TTGACAACAC
 31861 CGCAATAACA ATTAACGCTG GAAAAGGGTT GGAATTTGAA ACAGACTCCT CAAACGGAAA
 31921 TCCCATAAAA ACAAAATTTG GATCAGGACT ACAATATAAT ACCAATGGAG CTATGGTTGC
 31981 AAAACTTTGA ACAGGCTCA GTTTTGACAG CTCGGAGCC ATACAATGG GCAGCATAAA
 32041 CAATGACAGA CTTACTCTTT GGACAACACC AGACCCATCC CCAAAATGGA GAATTGCTTC
 32101 AGATAAAGAC TGCAAGCTAA CTCTGGCGCT AACAATAATGT GGCAGTCAAA TTTTGGGCAC
 32161 TGTTTCAGCT TTGGCAGTAT CAGGTAATAT GGCCTCCATC AATGGAAGCT TAAGCAGTGT
 32221 AAACCTTGGT CTTAGATTGT ATGACAAACG AGTGCTTATG TCAAAATTCAT CACTGGACAA
 32281 ACAGATTATG AACTTTAGAA ACGGGGACT CACTAACGGT CAACCATACA CTTATGCTGT
 32341 TGGGTTTATG CCAAACTTAA AAGCTTACCC AAAAATCTCAA AGTAAACTGA CAAAAGTAAA
 32401 TATTGTTAGC CAGGTGTATC TTAATGGTGA CAAGCTTAAA CCATTGCATT TTACTATTAC
 32461 GCTAATATGA ACAGATGAAA CCAACCAAGT AAGCAATAC TCAATAATAT CTAGTGGTCT
 32521 CTGGAACAGT GGACAAATCA CTAATGACAA ATTTGCCACC AATTCCPATA CCTTCCCTTA
 32581 CATTCGCCAG GAATAAGAAA TCGTGAACCT GTTGCAATGT ATGTTTCAAG GTGTTTATTT
 32641 TTCAATTTGA GAAAATTTCA AGTCATTTTT CATTCAGTAG TATAGCCOCA CCACCACATA
 32701 GCTTATACTA ATCACCGTAC CTTAATCAAA CTCACAGAAC CCTAGTATTC AACCTGCCAC

FIG. 7M

55/92

32761 CTCCTCCCA ACACACAGAG TACACAGTCC TTTCTCCCC GCTGCGCTTA AACAGCATCA
 32821 TATCATGGGT AACAGACATA TTCTTAGGTG TTATATTCCA CACGGTCTCC TGTCGAGCCA
 32881 AAGCTCATC AGTGATGTTA ATAACTCCC CGGGCAGCTC GCTTAAGTTC ATGTCCGCTGT
 32941 CCAGCTGCTG AGCCACAGGC TGCTGTCCAA CTTCGCGTTG CTCACGGGC GGCAGAGGAG
 33001 AAGTCCACGC CTACATGGGG GTAGAGTCAT AATCGTGCAT CAGGATAGGG CGGTGGTGTCT
 33061 GCAGCAGCGC GCGAATAAAC TGCTGCCGCC GCCGCTCGT CCTGCAGGRA TACAACATGG
 33121 CAGTGGTCTC CTCAGCGATG ATTGCGACCG CCGCAGCAT AAGGCGCCTT GTCTCCGGG
 33181 CACAGCAGCG CACCTTGATC TCACTTAAGT CAGCACAGTA ACTGCAGCAC AGTACCACAA
 33241 TATTGTTTAA AATCCACAG TGCAAGGCGC TGTATCCAAA GCTCATGGCG GGGACACAG
 33301 AACCACAGTG GCCATCATAC CACAAGCGCA GGTAGATTAA GTGCGACCC CTCATAAACA
 33361 CGCTGGACAT AAACATTACC TCTTTTGGCA TGTTGTAATT CACCACCTCC CGGTACCATA
 33421 TAAACCTCTG ATTAACATG GCGCCATCCA CCACCATCCT AAACCAGCTG GCCAAACCT
 33481 GCCCGCCGGC TATGCATGCG AGGGAACCGG GACTGGAACA ATGACAGTGG AGAGCCGAG
 33541 ACTCGTAACC ATGGATCATC ATGCTCGTCA TGTATCAAT GTTGGCACAA CACAGGCACA
 33601 CGCTGATACA CTCTCTCAGG ATTACAAGCT CCTCCCGGCT CAGAACCATA TCCAGAGGAA
 33661 CAACCCATTCT CTGAATCAGC GTAAATCCCA CACTGCAGGG AAGACCTGCG ACGTAACTCA
 33721 CGTTGTGTCAT TGTCAAAGTG TTACATTCCG GCAGCAGCGG ATGATCTCTC AGTATGTTAG
 33781 CGGTGTCTCT TGCTCAAAA GGAGGTAGGC GATCCCTACT GTACGGAGTG CCGCAGAGCA
 33841 ACCGAGATCG TGTTGGTCTG AGTGTCTATGC CAAATGGAAC GCCGAGCGTA GTCATATTTT
 33901 CTGAAGCAAA ACCAGGTGCG GCGGTGACAA ACAGATCTCG GTCTCCGGTC TCGTCCGTTA
 33961 GCTCGCTCTG TGTAGTAGTT TGTAGTATAT CACTCTCTCA AAGCATCCAG GCGCCCGCTG
 34021 GCTTCGGGTT CTATGTAAAC TCCTTCATGC GCGCTGCCCC TGATAACATC CACCACCGCA
 34081 GAATAAGCCA CACCAGCCA ACCTACACAT TCGTTCTGCG AGTCACACAC GGGAGGAGCG
 34141 GGAAGAGCTG GAAGAACCAT GTTTTTTTTT TTTATTCCAA AAGATTATCC AAAACCTCAA
 34201 AATGAAGATC TATTAAGTGA ACGCGCTCCC CTCGGGTGGC GTGGTCAAA TCTACAGCCA
 34261 AAGAAGACAT AATGGCAATT GTAAGATGTT GCACAAATGG TTCCAAAAGG CAAACTGCC
 34321 TCACGTCCAA GTGGACGTAA AGGCTAAACC CTTCAGGGTG AATCTCTCT ATAAACATTCT
 34381 CAGCACCTTC AACCATGCCC AAATAATTTT CATCTCGCCA CCTTATCAAT ATGTCTCTAA
 34441 GCAATCTCCG AATATTAAAG CCGGCCATTG TAAAAATCTG CTCCAGAGCG CCTCTCACTT
 34501 TCAGCCTCAA GCAGCGAATC ATGATTGCAA AAATTCAAGT TCCTCACAGA CTTGTATTAG
 34561 ATTCAAAAGC GGAACATTAA CAAAATACCC GCGATCCCGT AGGTCCCTTC GCAGGGCCAG
 34621 CTGAACATAA TCGTGCAGGT CTGCACGGAC CAGCGCGGCC ACTTCCCGCG CAGGAACCAT
 34681 GACAAAAGAA CCCACACTGA TTATGACACG CATACTCGGA GCTATGCTAA CCAGCGTACG
 34741 CCGGATGTAA GCTTGTGCA TGGGGCGGCA TATAAATGC AAGGTACTCG TCAAAAATTC
 34801 AGGCAAGGCC TCGCGCAAAA AAGCAAGCAC ATCGTAGTCA TGCTCAATGA GATAAAGGCA
 34861 GTTAAGTTCC GGAACACCA CAGAAAAGA CACCATTTTT CTCTCAAACA TGTTGCGGG
 34921 TTCTCTGATA AACACAAAAT AAAATAACAA AAAAAAATA ACATTAAAC ATTAGAAGCC
 34981 TGTNTTACAA CAGGAAAAC AACCCTTATA AGCATAAGAC GGACTAGCGC CATGCGGGG
 35041 TGACCGTAAA AAAACTGGTC ACCGTGATTA AAAAGCACCA CCGACAGTTC CTCGGTCATG
 35101 TCCGAGATCA TAATGTAAAG CTCGGTAAAC ACATCAGGTT GGTAAACATC GGTCAAGTGT
 35161 AAAAAGCGAC CGAAATAGCC CGGGGGAATA CATACCCGCA GCGGTAGAGA CAACATTACA
 35221 GCCCCCATAG GAGGTATAAC AAAATTAATA GGAGAGAAAA ACACATAAAC ACCTGAAAAA

FIG. 7N

56/92

35281 CCCTCCTGCC TAGGCAAAAT AGCACCCCTCC CGCTCCAGAA CAACATACAG CGCTTCACAA
35341 GCGGCAGCCA TAACAGTCAG CTTTACCAGT AAAAAACCT ATTAAAAAC ACCACTCGAC
35401 ACGGCACCAG CTCAATCAGT CACAGTGTAA AAAGGGCCAA GTACAGAGCG AGTATATATA
35461 GGACTAAAAA ATGACGTAAAC GGTAAAGTC CACAAAAACC ACCAGAAAA CCGCACGCGA
35521 ACCTACGCC AGAAACGAAA GCCAAAAAC GCACAACTTC CTCAAATCTT CACTTCCGTT
35581 TTCCCACGAT ACGTCACTTC CCATTTTAAA AAAAACTAC AATTCCTCAAT ACATGCCAAGT
35641 TACTCCGCC TAAACCTAC GTCACCGCC CGTTCCAC GCCCGCGCC ACGTCACAAA
35701 CTCACCCCC TCATTATCAT ATTGGCTTCA ATCAAATA AGGTATATTA TTGATGATG

FIG. 70

57/92

1 CATCATCAAT AATATACCTT ATTTTGGATT GAAGCCAATA TGATAATGAG GGGGTGGAGT
 61 TTOTGACGTG CGCGGGGGCG TGGGAACAAGG CGGGGTGAGC TAGTAGTGTG CGGGAATGTG
 121 GATGTTCGAA GTGTGGCGGA ACACATGTAA CGGACGGATG TGGCAAAAGT GAGCTTTTGTG
 181 GTGTGGCGCG GTGTACACAG GAAGTGACAA TTTTCGCGCG GTTTTAGCGC GATGTGTGTG
 241 TAAATTTTGG CGTAACCCAG TAAGATTTTG CCAATTTTCC GGGAAAACTG AATTAAGAGA
 301 AGTGAATCTT GAATAATTTT GTGTACTACT TAGCGCGTAA TAGCGCGTAA TATTTGTCTA
 361 GACTTTTGAAC GTTTACGTGG AGACTCGCCC AGGTGTTTTT CTCAGGTGTG TTCGCGGTTC
 421 CGGGTCAAAG TTGGCGTTTT ATTATTATAG TCAGCTGAGC AGTTTCTCTC TATACCGAAG
 481 TGAGTTTCTC AAGAGGCCAC TCTTGAGTGC CAGCGAGTAG AGTTTCTCTC TCCGAGCCGC
 541 TCCGACCCGC GGACTGAAAA TGAGACATAT TATCTGCCAC GGAGGTGTTA TATCCGAAGA
 601 AATGGCCGCC AGTCTTTTGG ACCACGTGAT CGAAGAGGTA CTGGCTGATA ATCTTTCACC
 661 TCCTAGCCAT TTGAAACCAC CTACCTCTCA CGAACTGTAT GATTTAGAGC TGACGGCCCC
 721 CGAAGATCCC AACGAGGAGG CGGTTTCGCA GATTTTTCCT GACTCTGTAA TGTGTGGCGT
 781 GCAGGAAGGG ATTGACTTAC TCACTTTTCG GCCCGGCCCG GGTCTCTCCG AGCCGCCCTCA
 841 CCTTTCCTCG CAGCCCGAGC AGCCCGGAGA GAGAGCCTTG GGTCCGGTTT CTATGCGAAA
 901 CCTTTCCTCG GAGGTGATCG ATCTTACCTG CCACGAGGCT GCCTTTCAC CCAAGTAGCA
 961 CGAGGATGAA GAGGGTGAGG AGTTTGTGTT AGATTATGTG GAGCACCCCG GGCACGGTGT
 1021 CAGGTCTGTG CATTAATCAC GGAGGAATAC GGGGGAACCA GATATTATGT GTTCGCTTTG
 1081 CTATATGAGG ACCTGTGGCA TGTTTGTCTA CAGTAAGTGA AATATTGGG CAGTGGGTGA
 1141 TAGAGTGTGT GGTTTGTGTG GTTAAATTTT TTTTAAATTT TACAGTTTTT GTGGTTTAAA
 1201 GAATTTTGTG TTGTGATTTT TTTAAAGGTG CTTGTGCTCG AACCTGAGCC TGACTTCGCG
 1261 CCAGAACCCG AGCCTGCAAG ACCTACCCGC CTCTCTAAAA TGGCGCCTGC TATCTCGAGT
 1321 GCGCCGACAT CACCTGTGTC TAGAGAATGC AATAGTAGTA CGGATAGCTG TGACTTCGCG
 1381 CTTCTTAACA CACCTCTCTG GATACACCCG GTGGTCCCGC TGTGCCCATT TAAACCAAGT
 1441 GCGGTGAGAG TTGGTGGCGC TCGCCAGGCT GTGAATGTA TCGAGGACTT GCTTAAACAG
 1501 CTTGGGCAAC CTTTGGACTT GAGCTGTAAA CGCCCGAGGC CATAAAGGTG AAACCTGTGA
 1561 TTGCTGTGTG GGTAAACGCC TTTTGTGCT GAATGAGTTG ATGTAAGTTT AATAAGGGT
 1621 GAGATAATGT TTAATCTGCA TGGCGTGTTA AATGGGCGG GCCTTAAAGG GTATATAATG
 1681 GCGCGTGGGC TAATCTTGCT TACATCTGAC CTCTGGAGG CTTGGGAGTG TTTGGAAGAT
 1741 TTTCTGCTGT TGCCTAACTT GCTGGAACAG AGCTCTAACA GTACCTCTTG GTTTTGGAGG
 1801 TTTCTGTGGG GCTCATCCCA GGCAAACTTA GTCTGCAGAA TTAAGGAGGA TTAAGATGG
 1861 GAATTTGAAG AGCTTTTGAA ATCTGTGTGT GAGCTGTTTG ATCTTTTGAA TCTGGGTAC
 1921 CAGGCGCTTT TCCAAGAGAA GGTCACTCAAG ACTTTGGATT TTTCCACACC GGGCGCGCT
 1981 GCGGCTGCTG TTGCTTTTTT GAGTTTATA AAGGATAAAT GGAGCAAGA AACCCATCTG
 2041 AGCGGGGGGT ACCTGCTGGA TTTTCTGGCC ATGCATCTGT GGAGAGCGGT TGTGAGACAC
 2101 AAGATCTGCC TGCTACTGTT GTCTTCCGCT CGCCCGGCGA TAATACCGAC GAGGAGGACG
 2161 CAGCAGCAGC AGGAGGAAGC CAGGCGGCGG CGGACGAGGC AGAGCCCATG GAACCGGAGA
 2221 GCGGCGCTGG ACCCTCGGGA ATGAATGTGT TACAGGTGGC TGAATCTGAT CCAAGACTGA
 2301 GAGCATATTT GACAATTACA GAGGATGGGC AGGGGTAAA GGGGGTAAAG AGGGAAGCGG
 2341 GGCTTTGTGA GGCTACAGAG GAGGCTAGGA ACTTAGCTTT TAGCTTAATG ACAGGACACG
 2401 GCTCTGAGTG TATTACTTTT CAACAGATCA AGGATAATTG CGCTAATGAG CTTGACTCTG
 2461 TGCGCGAGAA GTATTCCATA GAGCAGCTGA CCACTTACTG GCTGCAGCCA GGGGATGATT
 2521 TTGAGGAGGA TATTAGGTA TATGCAAAAG TGGCACTTAG GCCAGATTGC AAGTACAAGA
 2581 TCAGCAAACT TGTAAATATC AGGAATTTGT GATCAATTTT TGGGAACCGG GCGGAGGTG
 2641 AGATAGATAC GGAGGATAGG GTGGCTTTTA TCTCACAGG TGTAAAGCTC TATGGGTTTA
 2701 TGCTTGCATG GGACGGGGTG GTTATTATGA ATTAAGGTT TACTGGCCCC AATTTTAGCG
 2761 GTACGCTTTT CCTTGCCAAAT ACCAACCTTA TCTCACAGG TGTAAAGCTC TATGGGTTTA
 2821 ACAATACCTG TGTGGAGGCC TGGACCGATG TAAAGGTTCT GGGCTGTGCC TTTTACTGCT
 2881 CAGTGAAGGG GGTGTGTGTG CQCCCCAAA CAGGGCTCTC AATTAAGAA TGCTCTTTG
 2941 AAAGGTGTAC CTTGGGTATC CTGCTCGAGG GTAACCTCAG GTTGCGCCAC AATGTGGCTC
 3001 CGCACTGTGG TTGCTTCTAT CTAGTGAAA CCGTGGCTCT GATTAAAGAT AACATGTGAT
 3061 GTGGCAACTG CGGAGGACGG GCTCTCTAGA TGCTGACCTG CTCGGACGCG AACTGTACC
 3121 TGCTGAAGAC CATTCACGTA GCCAGCCACT CTGCAAGAGC CTGGCCAGTG TTAGAGATA
 3181 ACATACTGAC CCGCTGTTCC TTGCAATTTG TTTGAGGAG GGGGGTCTTC TACTCTTACC
 3241 AATGCAATTT GAGTCACACT AAGATATTGC TTGAGCCCGA GAGCATGTCC AAGGTGAACC

FIG. 8A

58/92

3301. TGAACGGGGT GTTTGACATG ACCATGAAGA TCTGGAAGGT GCTGAGGTAC GATGAGACCC
 3361. GCACCAGGTG CAGACCTGCG GAGTGTGGCG GTAACATAT TAGGAACACG CCTGTGATGC
 3421. TGGATGTGAC CGAGGAGCTG AGGCCGCGATC ACTTGGTGTCT GGCCTGCACC CGCCTGAGT
 3481. TTGGCTCTTAG CGATGAAGAT ACAGATTGAG GTACTGAAAT GTGTGGCGCT GGCCTTAAGGG
 3501. TGGGAAGAA TATATAAGGT GGGGGTCTTA TGTAGTTTGT TATCTGTTTT CGACGACCGC
 3641. CCGCGCCCAT GAGCACCAAC TCGTTTGTAG GAAGCATTTG GAGCTCATAT TTGACAAACG
 3661. GCATGCCCCC ATGGGCCGGG GTGCGTCAGA ATGTGATGGG CTCACGCAIT GATGGTCGCC
 3721. CCCTCTCGCC GCCTAACTCT ACTACTTTGA CTTACGAGAC CGTGTCTGGA ACGCGGTGG
 3781. AGACTGCAGC CTCGCCGCC GCTTCAGCGC CTGACGCCAC CGCCCGCGGG ATTGTGACTG
 3841. ACTTTGCTTT CCTGAGCCCG CTTCGAAGCA GTGCAGCTTC CGGTTCAATC GCCCGGATG
 3901. ACAAGTTGAC GCTCTTTTG GCACAAITGG ATTCTTTGAC CGGGGAACIT AATGTGCTTT
 3961. CTCAGCAGCT GTTGATCTG CGCCAGCAGG TTTCTGCCTT GAAGGCTTCT TCCCTCCCA
 4021. ATGCGGTGTA AAACATAAAT AAAAACCAG ACTCTGTTTG GATTTGGATC AAGAAGTGT
 4081. CTTCTCTGCT TTATTTAGGG GTTTTGCGCG CGCGGTAGGC CGGGACACAG CGGCTCCGT
 4141. CTTGTAGGGT CCTGTGTATT TTTTCCAGGA CGTGTGTAAG GTGACTCTGG ATGTTCAAGT
 4201. ACATGGGCAT AAGCCCTCT CTGGGGTGGG GTAGCACCA CTGCAGAGCT TCATGCTGCG
 4261. GGGTGGTGT GTAGATGATC CAGTCTGATC AGGAGCGCTG GCGCTGGTGC CTAATAATGT
 4321. CTTTCAGTAG CAGCTGATT GCCAGGGGCA GGCCTTGGT GTAAGTGTGT ACAAGCGGT
 4381. TAAGCTGGGA TGGGTGCATA CGTGGGGATA TSGATGATCT CTTGGAGCTGT ATTTTATAGT
 4441. TGCTATGTT CCCAGCCATA TCCCTCCGGG GATTCATGTT GTGCAAGAC ACCACGACAG
 4501. TGATATCGGT GCACCTGGGA AATTTGTGAT AGTCTTAGA AGGAAATGCG TGGGAAGTCT
 4561. TGGAGACGCC CTTGTGACCT CCAAGATTTT CCATGCAATC GTCCATTAAT ATGGCAATGG
 4621. GCCCGCGGGC GCGCGCTGG GCGAAGATAT TCTGGGATC ACTAACGTCA TAGTTGCTGT
 4681. CCAGGATGAG ATCGTCATAG GCCATTTTTT CAAAGCGCGG GCGGAGGGTG CCAGACTGCG
 4741. GTATATAGTT TCCATCCGGC CCAGGGGCGT AGTTACCTCT ACAGATTGAC ATTTCCCGT
 4801. CTTTGAGTTC AGATGGGGG ATCATGTCTA CCGTGGGGG GATGAAGAAA ACGGTTTTCG
 4861. GGGTAGGGGA GATCAGCTGG GAAGAAAGCA GGTTCCTGAG CAGCTGCGAC TTACCGCAGC
 4921. CGGTGGGGCC GTAATACACA CCTATTACCG GGTGCAACTG GTAGTTAAGA GAGCTGCAGC
 4981. TGCCGTATC CTTGAGCAGG GGGGCGCACT GTTAAAGCAT GTCCCTGACT CGCATGTGTT
 5041. CCCTGACCAA ATCCGCCAGA AGGCCGCTCG CGCCACGCGA TAGCAGTTCT TGCAGGAAG
 5101. CAAAGTTTTT CAACGGTTTG AGACCGTCCG CGGTAGGCAT GCTTTTGAGC GTTTGACCAA
 5161. CGAGTTCACG GCGGTCACC AGCTCGGTCA CCTGCTCTAC GGCATCTCGA TCCGACATAT
 5221. CTCTCTGTTT CGCGGGTTGG GCGGCTTTC GCTGTACGCG AGTAGTCGGT CCGTCTCAG
 5281. ACGGGCCAGG GTCATGTCTT TCCACGGGCG CAGGGTCTCT GTCAGCGTAG TCTGGGTCTG
 5341. GGTGAAGGGG TGCGCTCCGG GCTGCGCGCT GGCACAGGGT CGCTTGAAGC TGTCTTCAC
 5401. GGTGCTGAAG CGCTGCCGGT CTTTGCCTCG CGCGTGCGCC AGGTAGCATT TGACCATGTT
 5461. GTCATATAGC AGCCCTCCCG CGCGGTGGCC CTTGGCGGCG AGCTTGCCCT TGGAGGAGTC
 5521. GCGCAGCAGG GGGCAGTGCA GACTTTTGA GCGCTAGAGC TTGGGCGCGA GAAATACCGA
 5581. TTCCGGGGAG TAGGCATCCG CGCCGAGGCG CCGCGAGAGC GTCTCGCATT CCACGAGCCA
 5641. GGTGAGCTCT GCGCGTTCGG GGTCAAAAC CAGGTTTCCC CCAATGCTTT TGAATGCTTT
 5701. CTTACTCTG GTTCCATGA CGCGGTGTC ACCTCGGTTG ACGAAAAGCG TGTCGCTTAC
 5761. CCGTATGACA GACTTGAGAG GCTGTCTCT GAGCGGTGTT CCGGGTCTCT CCTGTATATG
 5821. ARACTCCGAC CACTCTGAGA CAAAGGCTCG AGCAGAGCC AGCAGGAAG AGGCTTAAGT
 5881. GGAGGGGGTAG CGGTGTTGTT CCACTAGGGG GTCCACTCGC TCCAGGGTGT GAAGACATAT
 5941. GTGCCCTCT TCGGCATCAA GGAAGGTGAT TGGTTTGTAG GTGTAGGCGA CCGTACCGGG
 6001. TGTTCCTGAA GGGGGGCTAT AAAAGGGGGT GGGGGCGGCT TCGTCTTAC TCTCTTCCG
 6061. ATGCTGTCT CGAGGGGCCA GCTGTTGGGG TGAAGTACTCC CTCTGAAGA CGGGCATGAC
 6121. TCTCTGCTAA AGATTGTGAG TTTCCAAAA CAGAGGAGAT TTGATATTCA CCTGGCCCGC
 6181. GGTGATCCCT TTGAGGGTGT CCGCATCTCC CTGTGACAAA AAGACAATCT TTTTGTGTC
 6241. AAGCTTGGTG GCAACAGACC CGTAGAGGCG GTTGACACG AACCTTGGCGA GTGAGCGCAG
 6301. GGTTGGTCTT TTGTGCGGAT CGCGCGGCTC GTGCGCGC ATGTTTAGCT GCACGTATTC
 6361. GCGCGCAACG CACCGCAATT CCGGAAAAGC GGTGGTGGCG TCGTCGGGCA CCGAGTGCAC
 6421. CCGCCAAACG CGGTTGTGCA GGTGTACAG GTCAACGCTG GTGGCTACT CTCCGCTAGG
 6481. CGGCTCGTTG GTCCAGCAGA GCGGCGCCCT CTTGCGCGAG CAGAATGGCG GTAGGGGGTC
 6541. TAGCTCGGTC TCGTCCGGGG GGTCTGCGTC CACGGTAAAG ACCCCGGGCA GCAGCGCGCG

FIG. 8B

59/92

6601 GTGGAAGTAG TCTATCTTGC ATCCTTGCAA GTCTAGCGCC TGCTGCCATG CGCGGGCGGC
 6661 AAGCGCGCCG TCGTATGGGT TGAAGTGGGG ACCCATCGGC ATGGGGTGGG TGAAGCGGGA
 6721 GGGGTACATG CCGCAAAATG CTGTAAGGTA GAGGGGCTCT CTGAGTATTC CAAGAATATG
 6781 AGGGTAGCAT CTTCACCGC GGAATCTGCG GCACAGGTAA TCGTATAGTT CGTGACGAGT
 6841 AGCGAGAGGG TCGGACCGA GGTTCGTACG GCGGGGCTCG TCTGCTCGGA AGACTATCTG
 6901 CCTGAAGATG GCATGTGAGT TGGATGATAT GGAATGATAT GGAAGCGTAG GAGTCGCGCA GCTTGTGTAG
 6961 GTCTGTGAGA CCTACCGCGT CACGCACGAA GCAGTAGTCC AGGGTTTCTT TGATGTGATC
 7021 CAGCTCGCGG GTGACCTGCA CGTCTAGGCG GCAGTAGTCC AGGGTTTCTT TGATGTGATC
 7081 ATACTTATCC TGTCCCTTTT TTTTCCACAG CTCGGCGTTG AGGACAAACT CTTCGCGGTC
 7141 TTTCAGTACG TCTTGGATCG GAAACCCGCT GGCCTCCGAA CGGTAAAGAG CTGATGATGA
 7201 GAATCTGGTG ACGGCTGGT AGGCACGACA TCCCTTTTCT ACGGGTAGCG CTAATGCGTG
 7261 CGCGGCTCTG CGGAGCGAGG TGTGGGTGAG CGCAAAAGGTG TCCCTGACCA TGACTTTGAG
 7321 GTACTGGTAT TTGAAGTCAG TGTCTGCGCA TCCGCCCTGC TCCAGAGCA AAAAGTCCGT
 7381 GCGCTTTTGT GAACGCGGAT TTGGCAGGGC GAAGGTGACA TCGTTGAAGA GTATCTTTCC
 7441 CGCGCAGGCG ATAAAGTTGC GTGTGATGCG GAAGGGTCCC GGCACCTCGG AACGTTTGTG
 7501 AATTACCTGG CGGCGGAGCA CGATCTCGTC AAAGCGGTTG ATGTTGTGCG CCACAAATGA
 7561 AAGTCCCAAG AAGCGCGGGA TGCCCTTGAT GGAAGCAAT TTTTAAAGT CCTCGTAGGT
 7621 GAAGCTCTCA GGGGAGCTGA GCCCTGTGCT TGAAGGGGCC CAGTCTGCAA GATGAGGGTT
 7681 GGAAGCGAGC AATGAGCTCC ACAGGTCACG GGCCTATGAG ATTTGCAAGT GTGCGCGGAA
 7741 GGTCTTAAAC TGGCGACCTA TGGCCATTCT TCTGGGGTGT ATGCAATAGA AGGTAACGG
 7801 GTCTGTGTTC CAGCGGTCCC ATCCAAAGTT CCGGGCTAGG TCTCGCGCGG CAGTCAGCAT
 7861 AGGCTCATCT CCGCGGAATC TCATGACCA GATGAAGGGC ACAGAGCTGT TCCCAAAGGC
 7921 CCCATCCAA GTATAGGTCT CTACATCGTA GGTGACAAAG AGACGCTCGG TGCGAGGATG
 7981 CGAGCCGATG GGAAGAATC GGAATCCTCC CCACCAATTT GAGGAGTGGC TATGTATGTG
 8041 GTGAAGTAG AAGTCCCTGC GACGCGGCCA ACACCTGTGC TGGCTTTTGT AAAAACGTGC
 8101 GCAGTACTGG CAGCGGTGCA CCGGCTGTAC ATCCTGACAG AGGTTGACCT GAGCAGCGCG
 8161 CACAAGGAAG CAGAGTGGGA ATTTGAGCCC CTCGCTGCG GGGTTTGGCT GGTGTCTTTC
 8221 TACTTCGGCT GCTTGTCTT GACCGTCTGG CTGCTCGAGG GGAATTAACG TGAATCGGAC
 8281 CACCACGCG CGCGAGCCCA AAGTCCAGAT GTCCGCGCG GCGGGTCGGA GCTTGTATGC
 8341 AACATCGCGC AGATGGGAGC TGTCCATGTT CTGGAAGTCC CCGGCGGTCA GGTCAAGCGG
 8401 GAGCTCTTGC AGGTTTACCT CGCATAGAGC GGTCAAGGCG CGGGCTAGAT CCAGGTGATA
 8461 CTAATTTTCC AGGGGCTGGT TGGTGGCGGC GTGATGGCT TCGAAGAGGC CGCATCCCG
 8521 CGCGCGACT ACGGTACCGC GCGGCGGCGC GTGGGCGCG GGGGTGTCTT TGGATGATGC
 8581 ATCTAAAGC GGTGACGCG GCGAGCCCC GAGGTAGGG GGGGCTCCG ACCCGCCGG
 8641 AGAGGGGGCA GGGCACGTC GCGCGCCGCG CGGGCAGGA GCTGTGCTG CGCGCTAGG
 8701 TTGCTGGCGA ACGCGACGAC GCGGCGGTTG ATCTCTGAA TCTGGCGCT CTGCTGGAAG
 8761 ACAGCGGGCG CGGTGAGCTT GAGGCTGAAA GAGAGTTGCA CAGAATCAAT TTCGTTGCTG
 8821 TTGACGGCGC CTGCGCGAA AATCTCTCTC ACGTCTCTGT AGTTGTCTGT ATAGGGGATC
 8881 TCGGCGCAGA ACTGCTGAT CTCTTCTCG TGGAGATCT CCGCTCGCG CATGACCAC
 8941 GTGGCGCGCA GGTCTGTGGA AATGCGGGCC ATGAGCTGCG AGAAGCGGTT GAGGCTTCCC
 9001 TGTGTTCCAGA CCGCGCTGTA GACCAACGCC CTTTGGCAT CCGGGCGCG CATGACCAC
 9061 TGCGGAGATG TGAAGTCCAC GTGCGGGCG AGACGCGCT AGTTTCTGAG CGCTGAAAG
 9121 AGGTAGTTGA GGTGGTGGC GGTGTGTTCT GCCACAAGA AGTACATAAC CCAGCTGTCG
 9181 AAGCTGGAAT GTTGATATC CCCCAGGCC TCAAGGCGCT CCATGGCTTC GTAGAGTCTC
 9241 GCGGGAATG TGAAGAACTG GAGGTGCGC GCGGACAGG TTAACCTCTT CTCACAGAAG
 9301 CGAGTGAAGT CCGCGACAGT GTGCGCACCC TCCGCTCTAA AGCTACAGG CGCTCTTCTT
 9361 TCTTCTTCTA TCTCTCTTTC CATAGGGGCC TCCCTTCTTT CTTCTCTGCG GCGCGGTGGG
 9421 GAGAGGGGGA CAGCGCGGCG ACGACGCGCG ACCGGAGGCG GGTGCAAGAA CGCTGCGATC
 9481 ATCTCCCGCG GCGACGCGC CATGCTCTCG GTGACGCGCG GCGGCTTCTC GCGGCGCGC
 9541 AGTTGGAAGA CGCGCGCGGT CATGTCCCGG TTAGGGTTTG CGGGGGGCTT GCGATGCGC
 9601 AGGATACCGC CGCTAACGAT CATCTCAAC AATTGTTGTG TAGGTACTCC CGCGCGAGG
 9661 GAGCTGAGCG AGTCCGCATC GACCGGATCG GAAACCTCTT CGAAGAAAGC GTCTAACAGG
 9721 TCACAGTCCG AAGGTAGGCT GAGCACCCTG CCGGGCGGCA CGGGCGGGCG GTGCGGGTTG
 9781 TTTCTGCGCG AGGTGCTGCT GATGATGTAA TTAAGGTGGA GCGCTCTGAG ACGGCGGATG
 9841 GTGACAGGAA GCACCATGTC CTTGGGTCG CCGTGTCTGA TGCGCAGGCG GTGCGCATG

FIG. 8C

60/92

9901 CCCGAGCCTT CGTTTGTACA TCGGCGCAGG TCITTTAGT AGTCITTCAT GAGCCTTTCT
 9961 ACCGCGACCT CTCTCTCTCC TTCTCTTTGT CCGTCATCTC TTGCATCTAT CGCTGGGGCG
 10021 GCGGCGGAGT TTGGCCGTAG GTGGCGCCCT CTCTCTCCCA TGCCTGTGAC CCCGAGCCCC
 10081 CTATCGCGCT GAAGCAGGGC TAGTTCGGCG ACAACGCGCT CGGCTAATAT GGCCTGTCTGC
 10141 ACCTGCGTGA GGGTAGACTG GAAATCATCC ATGTCACAAA AGCGGTGGTA TGCGCCGTG
 10201 TTGTGGTGT AAGTCAGATT GCCATAACG GCGCATTAAC CGGCTGTGGT ACCCGCTGCG
 10261 GAGAGCTCGG TGTAAGTGG AGCGGAGTAA GCCCTCGAGT CAAATACGTA GTCTGTGCAA
 10321 GTCGCGACCA GGTACTGGTA TCCACCAAAA AAGTGCGGCG GCGGCTGGCG GTAGAGGGCG
 10381 CAGCGTAGGG TGGCGGGGCG TCCGGGGGCG AGATCTTCCA ACATAAGGCG ATGATATCCG
 10441 TAGATGTACC TGGACATCCA GGTGATGCCG GCGGCGGTGG TGGAGGCGCG CGGAAAGTCG
 10501 CGGACGCGGT TCCAGATGTT GCGCAGCGCG AAAAAAGTCT CCATGGTCGG GACGCTCTGG
 10561 CGGTCAGGG GCGCGCAATC GTTGACGCTC TAGACCGTGC AAAAGGAGAG CTTGTAAGCG
 10621 GGCACTCTTC CGTGGTCTGG TGGATAAATT CGCAAGGGTA TCATGGCGGA CGACCGGGGT
 10681 TCGAGCCCGG TATCCGGCGC TCCGCCGTGA TCCATGCGGT TACCGCCCGC GTGTGCAACC
 10741 CAGGTGTGCG ACGTGAGACA ACGGGGGAGT GCTCCTTTTG GCTTCTTCC AGGCGCGCG
 10801 GCTGCTGCGC TAGCTTTTTT GGCCTACTGG CCGCGCGCAG GTAAGCGGTT AGGCTGGAAA
 10861 GCGAAGACAT TAAGTGGCTC GCTCCCTGTA GCCGGAGGGT TATTTTCCAA GGGTGTAGTC
 10921 GCGGAGACCC CGGTTGAGT CTCGGACCGG CGGAGTTCGC GCGAACCGGG GTTGTGACCG
 10981 CCGTCTGACA AGACCCCGCT TGCATAATCC TCCGGAAACA GGGACGAGCC CCTTTTGTGC
 11041 TTITTTCCAGA TGCATCCGGT GCTGCGCGAT ATGCGCCCCC CTCTTCAGCA CGGCGAAGAG
 11101 CAGAGCAGCG GGCAGACATG CAGGCGACCC TCCCTCTCTC CTACCGCGTC AGGAGGGGGCG
 11161 ACATCCCGGG TTGACGCGGC AGCAGATGGT GATTACGAAC CCCCAGCGCG CGGGCGAAGC
 11221 CACTACCTGG ACTTGGAGGA GGGCGAGGCG CTGGCGCGCG TAGGAGCGCC CTCTCCTGAG
 11281 CGGTACCCAA GGGTGCAGT GAAGCGTGAT ACCGCTGAGG CGTACGTGCC CGGCGAGAAC
 11341 CTGTTTCCGG ACCGCGAGGG AGAGGAGCCC GAGGAGATGC GGGATCGAAA GTTTCACGCA
 11401 GGGCGCGAGC TGGCGCATGG CCTGGAATCG GAGCGGTTTC TGC CGAGTTC GACATTTGAG
 11461 CCGGACGCGG GAACCGGGAT TAGTCCCGCG CGCGCACACG TGGCGGCCCG CGACCTGGTA
 11521 ACCGCATACG AGCAGACGGT GAACGAGGAG ATTAACTTTC AAAAAAGCTT TAACAAACAC
 11581 GTGCGTACGC TTGTGGCGCG CGAGGAGGTT GCTATAGGAC GCTATAGCAT GTGGCGAGCT
 11641 GTAAGCGCGC TGGAGCAAAA CCCAATATAG AAGCGCTCA TGGCGAGCT GTTCTCTATA
 11701 GTGCAGACAA GCAGGGACAA CGAGGCATTG AGGGATGCGC TGCTAAACAT AGTAGAGCCC
 11761 GAGGCGCGCT GGCTGCTCGA TTTGATAAAT ATCCTGCAGA GCATAGTGGT GCAGGAGCGC
 11821 AGCTTGTAGCC TGCTGACAAA GGTGGCGGCC ATCAACTATT CCATGCTTAG CCTGGGCGTT
 11881 TTTTACGCCC GCAAGATATA CCATACCCCT TACGTTCCCA TAGACAAGGA GGTAAAGATC
 11941 GAGGGGTTCT ACATGCGCAT GCGCGTGAAG GTGCTTACCT TGAGCGACGA CCTGGGCGTT
 12001 TATCGCAACG AGCGCATCCA CAAGGCCGTG AGCGTGAAGC GGCGGCGCGA GCTACGCGAC
 12061 CGCGAGCTGA TGCAAGCCTT GCAAAAGGCC CTGGCTGGCA CGGGCAGCGC GAGTAGAGAG
 12121 GCGAGTCTCT ACTTTGACGC GGGCGCTGAC CTGCGCTGGG CCCCAGCGCG ACGGCCCTTG
 12181 GAGGCGAGCT GGGCGGACCC TGGGCTGGCG GTGGCACCCG CGCGCGCTGG CAACGTCCGG
 12241 GGGCTGAGAG AATATGACGA GGACGATGAG TACGAGCCAG AGGACGGCGA GTACTATAGG
 12301 GTGATGTTTC TGATCAGATG ATGCAAGACG CAACGGACCC GCGGCTGCGG GCGGCGCTTC
 12361 AGAGCGACGC GTCCGGCCTT AACTCCACGG ACAGCTGGCG CCAAGTCAAT GACCGCATCA
 12421 TGTGCTGTAC TGGCGCAAT CTTGACGCGT CTTGACGAGA CCGCGAGGCC GCGCTGTGCG
 12481 CCGCAATCTT GGAAGCGGTG GTCCCGCGCG GCGCAACCC CACGACGAG AGGCTGTGCG
 12541 CGATCTGATAA CGCGCTGGCC GAAAAAGGG CCATCCGGCC CGACGAGGCC GAGCTGTGCT
 12601 ACAGACGGCT GCTTCAGCGC GTGGCTCGTT ACAACAGCGG CAACGTGCGA ACCAACCTGG
 12661 ACCGCTGGT GGGGGATGT GCGGAGGCGG TGGCGACGCG TGAGCGCGCG CACAGCTGAG
 12721 GCATCTGGG CTCCATGGTT GCACTAAACG CCTTCTGAG TACACAGGCC GGCACAGTGC
 12781 GCGGGGAGCA GGAGGACTAC ACCCACTTTG TGAGCGCATC CGGCTAATG CTTCAAGCTC
 12841 CACCGCAAG TGAGGTGTAC CAGTCTGGCG CAGACTATT TTTCCAGACC AGTAGACAAG
 12901 GCTCTGCGAC CGTAAACCTG AGCCAGGCTT TCAAAAACCT GCAGGGGCTG TGGGGGTGCG
 12961 GGGCTCCAC AGGCGACCGC GCGACGCTGT CTAGCTTGCT GACGCCCAAC TCGGCCCTAG
 13021 TGCTGCTGCT AATAGCGCCC TTCAACGGCA GTGGCAGCGT GTCCCGGAG ACATACCTAG
 13081 GTCACTTGCT GACACTGTAC GCGGAGGCCA TAGGTCAAGC GCATGTGGAC GAGCATACTT
 13141 TCCAGGAGAT TACAAGTGTG AGCCGCGCGC TGGGCGAGGA GGACACGGGC AGCCTGGAGG

FIG. 8D

61/92

13201 CAACCTAAA CTACCTGCTG ACCAACCGGC GGCAGAAAGT CCCCCTGGTTG CACAGTTTAA
 13261 ACAGCGAGGA GGAGCGCAT TTGCGCTACG TGCAGCAGAG CGTGAAGCTT AACCTGATGC
 13321 CGCAGCGGGT AACGCCACG GTGCGCGTGG ACATGACCAG CGCAGACATG GAACCGGGCA
 13381 TGATGCTC TC AACCGGCCG TTTATCAACC GCCTAATGGA CTACTTGCAT CGCGCGGCCG
 13441 CCGTGAACCC CGAGTATTTC ACCAATGCGCA TCTTGAACCC GCACATGGCTA CGCGCCCGCTA
 13501 OTTTCCTACAC CGCGGGATTTC GAGGTGCCCG AGGGTAACGA TGGATTCTCT TGGGACAGCA
 13561 TAGACGACAG CGTGTTTTCC CGCAACCCGC AGACCTGCTC AGAGTTTGCAA CCGCGCGAGC
 13621 AGSCAGAGGC GCGCTCCGGA AAGGAAAGCT TCCGAGGCC AAGCAGCTTG TCCGATGCG
 13681 GCGCTGCGGC CCCGCGGTCA GATGCTAGTA GCCCATTTCC AAGCTTGATA GGGTCTCTTA
 13741 CACGACATCG CACCACCCGC CGCGGCTTGC TGGGCGAGGA GGAGTACTTA AACCACTCGC
 13801 TGCTGCGAGG CAGCGCGGAA AAAAACCTGC CTCCGGCAT TCCCAACAAC GGGATAGAGA
 13861 GCCTAGTGGG CAAGATGAGT AGATGGAAGA CGTACGCGCA GGAGCAGCAG GACGTGCCAG
 13921 GCCCGCGCCC GCCACCCGT CGTCAAMGGC ACGACCGTCA CGCGGGTCTG GTGTGGGAGG
 13981 ACGATGACTC GGCAGACGAC AGCAGCGTCC TGGATTGGGG AGGGAGTGGC AACCCGTTTG
 14041 CGCACCTTCG CCCAGGCTG GGGAGAATGT TTTAAAAAAA AAAAAGCATG ATGCAAAATA
 14101 AAAACCTCAC CAAGCCATG GCACCGAGCG TTGGTTTTCT TGTATTCCCC TTAGTATGCG
 14161 GCGCGCGCGC ATGATGAGG AAGGTCTCTC TCCTCTCTAC GAGAGTGTGG TGAGCGCGGC
 14221 GCCAGTGGCG GCGCGCTGCG GTTCTCCCTT CGATGCTCCC CTGGACCCGC CGTTTGTGGC
 14281 TCGCGCGTAC TCGCGGCTA CCGGGGGGAG AACACGATC GCTTACTCTG AGTGGGACAT
 14341 CTTATTGACG ACCACCCGTG TGTACTGTGT GGACAAACAG TCAACGGATG TGGCATCCCT
 14401 GACACTACAG AACGACCACA GCAACTTTCT GACCAACGGT ATTCAAACAT ATGCATACAG
 14461 CCGGGGGGAG GCAAGCACAC AGACCATCAA TCTTGAAGAC CGGTGCGACT GGGGCGGCGA
 14521 CTGGAACACC ATCTGCGATA CCAACATGCC AAATGTGAAC GAGTTCACTG TTACCAATAA
 14581 GTTTAAGGCG CGGTGATGCG TGTCGCGCTT GCCTACTAAG GACAATCAGG TGGAGCTGAA
 14641 ATACGAGTGT GTGGAGTTCA CGCTGCCCGA GGGCACTAC TCCGAGACCA TGACCATAGA
 14701 CTTATGTAAC AACGCGATCG TGGAGCACTA CTTGAAGATG GGCAGACAGA ACGGGGTCTT
 14761 GGAAGCGGAC ATCGGGGTAA AGTTTGACAC AAGCGAAGCC TTCCATCCAG ACATCATTTT
 14821 CACTGCTCTT GTCATGCTTG GGGTATATAC AAACGAGCC TTCCATCCAG ACATCATTTT
 14881 GCTGCGAGGA TCGGGGTGG ACTTCAACCA CAGCCGCTG AGCAACTTGT TGGCATCCG
 14941 CAAGCGGCAA CCCTTCCAGG AGGGCTTTAG GATCACTAC GATGATCTGG AGGGTGTAA
 15001 CATTCCGACA CTGTTGGATG TGGAGCGCTA CCAGCGGAGC TTGAAAGATG ACACCGAACA
 15061 GGGCGGGGT GGCAGGAGCG GCAGCAACAG CAGTGGCAGC GGC CGGGAAG AGAAGCTCAA
 15121 CGCGCGAGCC CGGCAATGC AGCCGGTGA GGCATGAAC GATCATGCCA TTCGCGCCGT
 15181 CACCTTTGCC ACACGGGCTG AGGAGAAGCG CGCTGAGGCC GAAGCAGCGG CCGAAGCTCC
 15241 CGCCCGCCGT CGGCAACCCG AGGTCAGAA GCCTCAGAAG AAACCGGTGA TCAAAACGCG
 15301 GACAGGAGG AGCAAGAAAC GCAGTTACAA CCTAATAAGC AATGACAGCA CTTTCAACCA
 15361 GTTACCGACG TGGTACCTTG CATACAAC TAAGCCGCCCT CAGACCGGAA TCCGCTCATG
 15421 GACCTTGCTT TGCACCTCTG ACGTAACTCG CGGCTCGGAG CAGGTCTACT GGTCTGTGCC
 15481 AGCAATGATG CAAGACCCCG TGACCTTCCG CTCCACGCGC CAGATCAGCA ACTTTCCGCT
 15541 GGTGGGCGCC GAGCTGTGTC CGTGCACTC CAAGAGCTTC TACAACGACC AGGCCGCTTA
 15601 CTCCCACACT ATCCGCCAGT TTACTCTCTC GACCCAGCTG TTCAATCGTT TCCCGAGAA
 15661 CCAATTTTG GCGCGCCCGC CAGCCGCCAC CATCACACC GTCAAGTAAA ACGTTCCTGC
 15721 TCTACAGCAT CACGGGACGC TACCCTGCG CCAACGATC GGAGGAGTGC AGCAGTGAAC
 15781 CATTTACTGAC GCCAGACGCC GCACCTGCC CTACGTTTAC AAGGCCCTGG GCATAGTCTC
 15841 CGCGCGCGCT CTATCGAGCC GCACCTTTTG ACACAGCATG TCCATCTTTA TATCGCCOAG
 15901 CAATAACACA GGCTGGGGCC TGCGCTTCCC AAGCAAGATG TTTGGCGGGG CCAAGAGAGC
 15961 CTCCGACCAA CACCCAGTGC CGTGCAGCG GCACATGCG CGGCCCTGGA GCGCGCGACA
 16021 AC CGCGGCCG ACTGGGCGCA CCAACGTCGA TGACGCCATC GACCGCGGTG TGGAGGAGGC
 16081 GCGCAACTAC ACGCCACGC CGCCACAGT GTCCAAGTG GACCGGCTCA CACCGGCTGC
 16141 GGTGCGCGGA GCCCGCGCT ATGCTAATAA GAAGAGACGG CGGAGGCGCG TAGCACTGCG
 16201 CCACCGCCGC CGACCCGCA CTGCGGCCCA CTGCGGCCCA AGCGCGCGCG CGCGCCCTGC
 16261 ACGTGCACCC GCGCGACGCG CGGCGATGCG GCGCGCTGCA AGCGCTGACG CGGATATTGT
 16321 CACTGTGCC CCCAGGTCCA GCGGACGAGC GCGCGACGCA GCAGCGCGGA CACTTAGTGC
 16381 TATGACTCAG GGTGCGAGGG GCACCTGTGA TTGGGTGCGC GACTCGGTTA GCGGCTGCG
 16441 CGTGGCCGCG CGCACCCGCC CCCCGCGCAA CTAGATTGCA AGAAAAAATC ACTTAGACTC

FIG. 8E

62/92

```

16501 GTACTGTTGT ATGTATCCAG CGGCGCGCGC GCACAACGAA GCTATGTCCA AGCGCAAAAT
16561 CAAAGAAGAG ATGCTCCAGG TGACTCGGCC GGAGATCTAT GGCCCCCGCA AGAAGGAAGA
16621 GCAGGATTAC AAGCCCGGAA AGCTAAAGCG GGCTAAAAGG AAAAAGAAAG ATGATGATGA
16681 TGAACCTTGAC GACGAGGTGG AACTGCTGCA CGCTACCGCG CCCAGCGCAG GGGTAGCAGT
16741 GAAAGGTGCA CGCTAAAAAC GTGTTTGGCG ACCCGGCACC ACCGTAGTCT TTACGCCGGG
16801 TGAGCGCTCC ACCCGCACCT ACAAGCGCGT GTATGATGAG GTGTACGGCG ACGAGGACCT
16861 GCTTGAGCAG GCCAACGAGC GCCTCGGGGA GTTTGCCCTAC GGAAGCGCGC ATAAAGACAT
16921 GCTGCGGTG CGCTGGACG AGGGCAACCC AACACCTAGC CTAAGCCCGC TAACACTGCA
16981 GCAGGTGCTG CCCGCGCTTG CACCGTCCGA AGAAAAGCGC GGCCTAAGCG GCGAGTCTGG
17041 TGACTTGGCA CCCACCGTGC AGCTGATGGT ACCCAAGCGC CAGCGACTGG AAGATGTCTT
17101 GGAAAAAATG ACCGTGGAAC CTGGGCTGGA GCCCGAGGTC CGCGTGCAGC CAATCAAGCA
17161 GGTGGCGCCG GGACTGGGCG TGCAGACCGT GGACGTTTCA ATACCCCACT CAGATAGCAC
17221 CAGTATTGCC ACCGCGCACG AGGGCATGGA GACACAAACG TC0CCGGTGT CCTCAGCGGT
17281 GCGGATGCC GCGGTGCAGG CGGTGCTGCG GGCCCGGTCC AAGACCTCTA CGGAGGTGGA
17341 AACGGACCCG TGGATGTTTC GCGTTCACG CCCCGCGCGC CGCGCGGTTT CAGGAAGTA
17401 CGCGCGCCGC AGCGCGCTAC TGCCCGAATA TGCCCTACAT CCTTCCATTG CGCCTACCCC
17461 CGGCTTACGT GGCTACACCT ACCGCCCCAG AAGACGAGCA ACTACCCGAC GCGCAACACC
17521 CATGGAAAGC CGCCGCGCGC CTGCGCCGTG CCAGCCCGTG CTGGCCCCGA TTTCCGTGAG
17581 CAGGCGTGCT CGCGAAGGAG GCAGGACCCCT GGTGCTGCCA ACAGCGCGCT ACCACC0CAG
17641 CATCGTTTAA AAGCGGTCT TGTGTTTCT TGCAAGATG GCCCTCAGCT GCGCGCTCCG
17701 TTTCCCGGTG CGGGGATTCG GAGGAAGAA GTACCGTAGG AGGGGCAATG CGGCGCATG
17761 CTTGACGGCG GGCATCGGCT GTGCGACCA CGCGCGCGCG CGCGCGTCCG ACCGTCCAGT
17821 CGGCGCGGCT ATCTCGCCCT TCCTTATTCC ACTGATCGCC GCGCGGATTG GCGCGTGGC
17881 CGGAATGCTA TCCGTGGCCT TGCAGCGCCA GAGACACTGA TTAATAACAA GTTCAATGTA
17941 GAAAAATCAA AATAAAAAGT CTGGAACCTC ACGCTCGCTT GGTCTGTAA CTATTTTGTA
18001 GAATGGAAAG CATCAACTTT GCGTCTCTG CGCCCGGACA CGCTTCGCGC CGGTCTATGG
18061 GAACTTGGCA AGATATCGGC ACCAGCAATA TGAGCGGTGG CGCCTTCACG TGGGGCTCGC
18121 TGTGGAGCGG CATTAATAAT TCGGTTCCA CCCTTAAGAA CTATGGCAAG AAGGCCCTGA
18181 ACAGCAGCAC AGGCCAGATG CTGAGGATA AGTGTAAAG GCAAAATTTT CAACAAAAGG
18241 TGTGATAGTG CTTGGCTCTC GGCATTAGCG CTTGATCCCC CCGGCTGGA CCGGCACTGC
18301 AAAATAAGAT TAACAGTAAG CTTGATCCCC GCGCTTCCGT AGAGGAGCCT CACCGCGCGC
18361 TGGAGACAGT GTCTCCAGAG GGGCGTGGCG AAAAGCGTCC GCGCCCGCAC AGGGAAGAA
18421 CTCGTGTGAC GCAAATAGAC GAGCCTCCCT CGTACGAGGA GGCCTAAGG CCAAGCTCTC
18481 CCACCACCCG TCCCATCGCG CCCATGGCTA CCGGATGTCT GGGCGACGAC ACACCCGTAA
18541 CGCTGGACCT GCCTCCCGCC GCGGACACCC AGCAGAAACC TGTGCTGCA TGGCCCGAGT
18601 CCGTGTGTTT AACCGTCTCT AGCGCGCGGT CCCCTGCGCG CGCCGCGCAG GGTCCGCGAT
18661 CGTTGCGGCC CGTAGCCAGT GGCACACTGC AAGCACACTT TAACGTGTG TATGTGTGTC
18721 GGGTGCATC CTTGAAGCGC CGACGATGCT TCTGAATAGC TAACGTGTG TATGTGTGTC
18781 ATGTATGCGT CCATGTGCCC GCGCAGGAGG CTGCTGAGCG GCGCGCGCGC CGCTTCCAA
18841 GATGGCTACC CTTTCGATGA TGCCGCGAGT GTCTTACATG CACATCTCGG GCGCAGGAGC
18901 CTTGGAGTAC CTGAGCCCGG GCGTGTGTCG GTTTGCCCGG GCGCGCTACG CACGACGTGA
18961 CTTGAATAAC AAGTTTAGAA ACCCAACGST GTGTCAGCTG TGTGACCTGT GAGGATACTG
19021 TCCCAAGGCT GTGACGTCG GGTTCATCGT CGTACGAGGA GAGGATCTG CTGACATGTC
19081 CAAGCGCGGG TTCAACCTAG CTGTGGGTGA TAACCGTGTG CTGGACATG CTTCACGTA
19141 CTTTGAATGC CGCGCGCTGC TGGCAGGGG CCTACTTTT AAGCCCTACT ATGCAAAACC
19201 CTACAAACCC CTGGCTCCCA AGGGTGC0CC AAATCCTTGC GAATGGGATG AAGCTGTAC
19261 TGCTCTTGAA ATAAACCTAG AAGAAGAGGA CGATGACAC CAAGACGANG TAGACGAGCA
19321 AGCTGAGCAG CAAAACACTC ACGTATTTTG GCAGGCGCCT TATTCTGGTA TAAATATTAC
19381 AAGGAGGGGT ATTCAAATAG GTGTGGAAG CTAAACACCT AAATATGCGC ATGCAAAACC
19441 TCGAGCTGAA CCTCAAATAG GAGAATCTCA GTGGTAGCAA ACTGAAATTA ATCATGACAG
19501 TGACGAGATC CTTAAAAAGA CTACCCCAAT GAACCCATGT TACGGTTTCT ATGCAAAACC
19561 CACAAATGAA AATGGAGGGC AAGGCATTCT TGTAAAGCAA CAAATGGAAG AGCTAGAAGG
19621 TGAAGTGAA ATGCAATTTT TCTCAACTAC TGAGCGCACT GAGGCGAAT GTGATAACTT
19681 GACTTCCTAA GTGGTATTGT ACAGTGAAGA TGTAGATATA GAAACCCGAG ACACCTATAT
19741 TTCTTACATG CCCACTATTA AGGAAGGTAA CTCACGAGAA CTAATGGGCG AACATCTAT

```

FIG. 8F

63/92

19801 GCCAACAGG CCTAATTACA TTGCTTTTAG GGACAATTTT ATTGGTCTAA TGTATTACAA
 19861 CAGCAGCGGT AATATGGGTG TTCTGGCGGG CAAAGCATCG CAGTTGAATG CTGTTGTAGA
 19921 TTGTCAGAC AGAAACACAG AGCTTTCATA CCAGCTTTTG CTTGATTCCA TTGGTGTATG
 19981 AACCAGGTAC TTCTCTATGT GGAATCAGCG TTGTTGCAGC TATGATCCAG ATGTTGATGA
 20041 TATTGAAAAAT CATGGAACTG AAGATGAATC TCCAAATTAC TGCTTTCCAC TGGGAGGTGT
 20101 GATTTAATAA GAGACTCTTA CCAAGGTAAA ACCTAAAAAA GGTCAAGAAA ATGGATGGGA
 20161 AAAAGATGCT ACAGAAATTT CAGATAAAAA TGAATAAAGA GTTGAAATAA ATTTTGCAT
 20221 GGAAATCAAT CTAAATGCCA ACCTGTGGAG AAATTTTCCT TACTCCAACA TAGCGCTGTA
 20281 TTGCCCAGC AAGCTAAAGT ACAGTCTCTT CAACGTAAAA ATTTCTGATA ACCCAACAC
 20341 CTACAGCTAC ATGAACAAGC GAGTGGTGGC TCCCGGGTTA GTGAGCTGT ACATTAACCT
 20401 TGGAGCAGC TGTGCCCTTG ACTATATGGA CAACGTCAAC CCAATTTAAC ACCACCGCAA
 20461 TGTGCGCTCG CGTACCCTGT CAATGTTGCT GGGCAATGGT CGCTATGTC CTTTCCACAT
 20521 CAGGTGCGCT CAGAAGTCTT TTGCCATTA AAACCTTCCT CTCCGCGCG GCTCATACAT
 20581 CTACAGGTGG AACTTCAGGA AGGATGTAA CATGGTTCCT CAGAGCTCCC TAGGAAATGA
 20641 CCTAAGGGTT GACGGAGCCA CATTAAAGTT TGATAGCAAT TGCCCTTAAG CCACCTTCTT
 20701 CCCATGSGCC CACAACACCG CCTCCACGCT TGAGGCCATG CTTAGAAAGC ACACCAACGA
 20761 CCACTGCTTT AAGCACTATC TCTCCGCGCG CAACATGCTC TACCTATAC CGGCCAAGCG
 20821 TACCAGCTGT CCCATATCCA TCCCCTCCCG CAACCTGGCG GCTTTCCGCG CTTTGGCCTT
 20881 CAGCGGCTT AAGACTAAGG AAACCCCATC ACTGGGCTCG GGCTACGAGT GCATTATAC
 20941 CTACTCTGGC TCTATACCTT ACCTAGATGG AACCTTTTAC CTCACACACA CCTTTAAGAA
 21001 GGTGGCCATT ACCTTTGACT CTCTGTGCA CTGGCTGGCG AATGACCGCC TGCTTACCCC
 21061 CAACGAGTTT GAAATTAAGC GCTCAGTTGA CGGGGAGGGT TACAACGTTG CCCAGTGTAA
 21121 CATGACCAAA GACTGTTTCC TGGTACAAAT GCTAGCTAAC TACAACATTG GCTACAGGAA
 21181 CTCTATATA CCAGAGAGCT ACAAGGACCG CATGTACTCC TTCTTTAGAA ACTTCGAGC
 21241 CATGAGCGCT CAGTGTGTGG ATGATACTAA ATACAAGGAC TACCAACAGG TGCGCATCTC
 21301 ACACCAACAC AACCACTCTG GATTTGTTGG CTACCTTGCC CCCACCATG CGGAGGAGCA
 21361 GGCTTACCCT GCTAACTTCC CCTATCCGCT TATAGGCAAG ACCGCAAGT ACAGCATTTAC
 21421 CCAGAAAAAG TTTCTTTGGC ATCGCACCTT TTGCGCATC CCAATTCPCA GTACCTTTAT
 21481 GTCCATGGGC GCACTCAGAG ACCTGGGACA AAACCTTCTC TACGCCAACT CGGCCCAAGC
 21541 GCTAGACATG ACTTTTGAGG TGGATCCCAT GGACGAGCCC ACCCTTCTTT ATGTTTGTGT
 21601 TGAAGTCTTT GACGTGGTCC GTGTGACCGG GCGCGACCG GGCGTCATCG AAACCGTGT
 21661 CTTGCGCAGC CCTTCTCGG CCGGCAACGC CACAACATAA AGAAGCAAGC AACATCAACA
 21721 ACAGCTGCGC CATTGGGCTC CAGTGAGCAG GAAGTGAAG CCAATTTGCA AGATCTTGGT
 21781 TGTGGGCGAT ATTTTGTGGG CACCTATGAC AAGCGCTTTC CAGGCTTTGT TTCTCCACAC
 21841 AAGCTCGCT CGGCCATAGT CAATACGGCC GGTCCGAGAG CTGGGGCGCT ACACCTGGATG
 21901 GCCTTTGCTT GGAACCCGCA CTCAAAAACA TGCTACCTCT TTGAGCCCTT TGGCTTTTCT
 21961 GACCAGCGAC TCAAGCAGGT TTACAGTGTG GAGTACGAGT CACTCCGCG CCCTAGCGCC
 22021 ATGTGCTCTT CCCCGACCG CTGTATAAGC CTGGAAGAGT CCACCCAAAG CGTACAGGGG
 22081 CCCAACTCGG CGCGCTGTGG ACTATTCTGC TGCAATGTTT TCCACGCTTT TGCACACTGT
 22141 CCCAAACTC CCAATGATCA CAACCCCAAC ATGAACCTTA TTACCGGGGT ACCCAACTCC
 22201 ATGCTCAACA GTCCCCAGGT ACAGGCCAAC CTGCGTCGCA ACCAGGAACA GCTCTACAGC
 22261 TTCTGAGGAC GGCATCTGCC CTACTTCCGC AGCCACAGTG CGCAGATTAG GAGCGCACT
 22321 TTTTTTTGTG ACTTGAAAAA CATGTAAAAA TAATGTACTA GAGACACTTT CAATAAAGAG
 22381 AAATGCTTTT ATTTGTACAC TCTCGGGTGA TTAATTTACG CCACCCCTTG CGTCTGGCGC
 22441 GTTTAAAAAT CAAAGGGGTT TCTGCGCGCA TCGCTATGCG CCAGCTGGCA GGCACAGTTG
 22501 CGATACTGGT GTTTAGTGCT CCACTTAAAC TCAGGCACAA CCAATCCGCG CAGCTCGGTT
 22561 AATGTTTTCAC TCCACAGGCT CGGCAACCAT ACCAAGCGCT TTAGCAGGTC GGGCGCGATG
 22621 ATCTGAAGAT CGCAGTTGGG GCTCCGCGCC TGCAGCGCGC AGTTGCGATA CACAGGGTTG
 22681 CAGCAGCTGA ACATATCAG CGCCCGGTGG TGCACGCTGG CCAGCAGCTT CTGTGCGGAG
 22741 ATCAGATCCG CGTCCAGGTC CTCCGCGTTG CTCAGGCGCA ACGGAGTCAA CTTTGGTAGC
 22801 TGCTTTCCCA AAAAGGGCGC GTGCCAGCG GTTGAAGTTG ACTCGCACCG TAGTGCGATC
 22861 AAAAGGTGAC CGTGCCCGGT CTGGCGGCTA GGAATACAGC CTTGCAATAA AGCTTGTATC
 22921 TCTTAAAAAT CCACCTGAGC CTTTGCGGCT TCAGAGAAGA ACATGCCGGA AGACTTCCGC
 22981 GAAAACTGAT TGGCCGAGCA GCGCCGCTCG TGCACGCAGC ACCTTGGCTG GGTGTTGGAG
 23041 ATCTGCACCA CATTTCCGGC CCACCGGTTT TTCAGATCTT TGCGCTTGCT AGACTGCTCC

FIG. 8G

64/92

23101 TTCAGCGCGC GCTGCCCGTT TTCGCTCGTC ACATCCATTT CAATCAGCTG CTCCTTATTT
 23161 ATCATAATGC TTCCGTGTAG ACACCTTAAGC TCGCCTTGG TCTCAGCGCA GCGGTGCAGC
 23221 CACAACCGCG AGCCCGTGGG CTCGTGATGC TTGTAGGTCA CCTCTGCAAA CGACTGCAGG
 23281 TAGCGCTGCA GGAATCGCCC CATCATGCTC ACAAGGTTCT TGTTGCTGGT GAAGTGCAGC
 23341 TGAACCCGCG GGTGCTCCTC GTTCAGCCAG GTCTTGCATA CGGCCCGCAG AGCTTCCACT
 23401 TGGTCAGGCA GTAGTTTGAA GTTCGCTTTT AGATCGTTAT CCACGTGGTA CTTGTCCATC
 23461 AGCGCGCGCG CAGCCTCCAT GCCTTCTTCC CACGCAGACA CGATCGGCAC ACTCAGCGGG
 23521 TTATCACCAG TAATTTCACT TTCCGCTTGG CTGGGCTCTT CTTCTTCTCT TTGCGTCCGC
 23581 ATACCACGCG CCACTGGGCT GTCTTCATTG AGCCCGCGCA CTGTGCGCTT ACCTCCTTTG
 23641 CATGCTTTGA TTAGCACCAG TGGGTTGCTG AAACCCACCA TTTGTAGCGC CACATCTTCT
 23701 CTTTCTTCTT CGCTGTCCAC GATTACCTCT GTTGATGCGC GGCCTCTGGG CTTGGGAGAA
 23761 GGGCGCTTCT TTTTCTTCTT GGGCGCAATG GCCMAATCCG CCGCCGAGGT CGATGGCCCG
 23821 GGGCTGGGTG TGGCGGGCAC CAGCGCTCTT TGTGATGAGT CTTCTCTGTC CTCGACTCG
 23881 ATACGCCGCC TCATCCGCTT TTTTGGGGCG GCGCCGGGAG GCGGCGCGGA CGGGACGGG
 23941 GAGCAGCGT CCTCCATGGT TGGGGGAGCT CGCGCCGCAC CGCGTCCGCG CTCGGGGGTG
 24001 GTTTCGCGCT GCTCTCTTCT CCGACTTGCC ATTTCTTCTT CCTATAGCA GAAAAAGATC
 24061 ATCGAGTCAG TCGAGAAGAA GGACAGCGTA ACCGCCCTCT CTGAGTTCGC CACTCAGGCC
 24121 TTGACCGATC CGGCCAACGC GCCTACCACC TTCCCGCTCG AGGCACCCCT GCGTGGGAG
 24181 GAGGAAGTGA TTATCGAGCA GGACCCAGGT TTTGTAAAGC AAGACGACGA GGACCGCTCA
 24241 GTGAAACACG AGGATAAAAA GCAAGACGAG GACAACGAG AGGCAACGA GGAACACGTG
 24301 GGGCGGGGGG AGCAAAAGCA TGGCGACTAC CTAGATGTGG GAGACGACGT CTGTGTGAAG
 24361 CATCTGACG GCGAGTGGCG CATTATCTGC GAGCGCTTGC AAGAGCGCAG CGATGTGCC
 24421 CTGCGCATAG CGGATGTGAG CTTTGCCTAC GAACGCCACC TATTCTCACC CGCGTACCC
 24481 CCCAAACCGG AAGAAACCGG CACATGCGAG CCCAACCCGC GCCTCAACTT TATCCCGGTA
 24541 TTTGCCGTGC CAGAGGTGCT TGCCACTAT CACATCTTTT TCCAAACTGT CAAGATACCC
 24601 CTATCTTGCC GTGCCAACCG CAGCCGAGCG GACAAGCAG CGGAAATGA AAGTCACTCT
 24661 GTCAATCTGT ATATCGGCTC GCTCAACGAA GTGCCAAAAA TCTTTGAGGG TCTTGGACCT
 24721 GACGAGAAGC GCGCGGCAAA CGCTCTGCAA CAGGAAGAAA GCGAAATGA AAGTCACTCT
 24781 GGAGTGTGGT TGGAACTCGA GGTGTACAA CCGCGCTTAG CCGTACTAAA AGCAGCAGTC
 24841 GAGGTCAACC ACTTTGCTTA CCGCGCACTT AACCTPACCC CCAAGGTACT GAGCAGAGTC
 24901 ATGAGTGAGC TGATCGTGGC CCGTSCGAG CCCCTGGAGA GGGATGCAAA TTTGCAAGAA
 24961 CAACAGAGGG AGGGCTTACC CGCACTTGGC GACGAGCAG CCCCTGGAGA GGGATGCAAA
 25021 CGGAGCGCTC CCGACTTGA GAGGCGACGC AAACATAAGA TGCGCCAGT TGCTGCTTAC
 25081 GTGGAGCTTG AGTGCATGCA CGGTTTCTTT TACGTACCGC AGGCTTGCAA GATCTCCAC
 25141 GAAACATTGC ACTACACCTT TCGACAGGGC GCAAGTTTGC GCGTCAACCG CTTTGGCGAA
 25201 GTGGAGCTCT GCAACCTGGT CTCTACCTCT GGAATTTTGC GCTGACCCGC ACTACGTCCG CGACTGCGT
 25261 AACGTGCTTC ATTCTACGCT CAAGGGCGAG GCGCGCCGCG CACTAGCTGT GCTCAACAGC
 25321 TACTTATTTCT TATGCTACAC CTGGCAGAGC CTAAAGACTC GTTGGCGAGA CTTTGGCAGA GTGCTTGGAG
 25381 GAGTGCACCC TCAAGGAGCT GCGCGCGCAC CTGCGGAGCA TCAATTTCCC CGAACCGCTT CGAGAACTTG
 25441 GCCTTCAACG AGCGCTTCTG TGGCGGAGG TGTGCGAGAT TTGCGGCCCA CTTGCTGTGC ACTTCTTACG
 25561 AGGAACCTTA TCCTAGAGCG CTCAGGAATC CTGCGAATGC CCGCGAATGC CTTGCGCGCG TTTGGGGCCA CTTCTAGCT
 25621 GACTTTGTGC CATTAAAGTA CCGCGAATGC TCTGACATAA TGGAGAGCTT GAGCGGTGAC
 25681 CTGACGCTAG CCAACTACCT TGCTTACCAC CTGCTGCAAC CTTATGACCC CGCACCGCT CCGTGTTTGC
 25741 GGTCTACTTG AGTGTCACTG TCGCTGCAAC TCGCTGCACT TTGAGCTGCA GGTGCCCTCG
 25801 AATTGCGACG TGCTTAAAGA AAGTCAAAAT ATCGGTACCT TTGAGCTGCA GGTGCCCTCG
 25861 CTGACGAAA AGTCCGCGGC TCGGCGGTG AAACCTCACT CCGGGCTGTG GCGTCCAGCT
 25921 TACTTTCGCA AATTGTGACC TGAGGACTAC CAGCGCCAGC AGATTAGGTT TATCAGAGAC
 25981 CAATCCCGCC CGCCAAATGC GGAGCTTACC GCCTGCTACA TTACCCAGGG CCAATTTCTT
 26041 GGCCAAATGC AAGCCATCAA CAAGCCCGC CAAGAGTTTC TGCTACGAAA GGGAGCGGGG
 26101 GTTACTTTGG ACCCCAGTCC CGCGGAGGAG CTCACCCCAA TCCCCCGCG CCGCGAGCGT
 26161 TATCAGACCG AGCCCGGGGC CTTTGTCTTC CAGGATGGCA CCCAAAAGAA AGCTCAGCT
 26221 CGCCCGGCA CCAACGGAGC AGGAGGAATA CTGGGACAGT CAGGACAGAG AGGTTTGTGA
 26281 CGAGGAGGAG GAGGACATGA TGGAGAGCTA GGAGAGCCTA GACGAGGAAG CTTTCCAGGT
 26341 CGAAGAGGTG TCAGACGAAA CACCGTACCC CTCGGTGC A TTTCCCTCG CCGCGCCCA

FIG. 8H

65/92

26401 GAAATCGGCA ACCGGTTCCA GCATGGCTAC AACCCTCCGCT CCTCAGGCGC CGCGGGCACT
 26461 GCCCGTTCGC CGACCCCAACC GTAGATGGGA CACCCTGGGA ACCAGGGCGC GTAAAGTCCAA
 26521 CGAGCCGCCG CCGTTAAGCC AAGAGCAACA ACAGCGCCAA GGCTACCCTG CATGCGCGGG
 26581 GCACAAGAAC GGCATAGTTG CTTCGTTGCA AGACTGTGGG GGCAACATCT CCTTCGCCCG
 26641 CCCTTCTCTT CTCTACCATC ACGGCTGTGG CTTCGCCCGT AACATCTCTG ATTACTACCG
 26701 FTATCTCTTAC AGCCCATACT GCACGCGCGC CAGCGCGCAG CGGCAGCAAC GACAGCGGCA
 26761 CACAGAAGCA AAGCGCAGCG GATAGCAAGA CTCTGACAAA GCCCAAGAAA TCCACAGCGG
 26821 CGCGAGCAGC AGGAGGAGGA CGCTGTGCTC TGCGGCCCAA CGAACCCTGA TCGACCCGCG
 26881 AGCTTAGAAA CAGGATTTTT CCCACTCTGT ATGCTATAAT TCAACAGAGC AGGGCGCAAG
 26941 AACAGAGGCT GAAATATAAA AACAGTCTC AACAGTCCCT CACCOCGAGC TGCTGTATTC
 27001 ACAAAAGCGA AGATCAGCTT CGCGCAGCGC TGGAGAGACC GGAGGCTTTC TTCACTAAAT
 27061 ACTGCGCGCT GACTCTTAAG GACTAGTTTC CACCCGCGCC CAGCACCCTGT GGTACAGCGC
 27121 TACGTATCTT CCAGCGGCCA CACCCGCGCC CAGCACCCTGT GGTACAGCGC
 27181 AGGAAATATC CACGCCCTAC ATGTGGAAT ACCACGCCCA AATGGGACTT GCGGCTGGAG
 27241 CTGCCCAAGA CTACTCAACC CGAATAAACT ACATGAGCGC GGGAGCCCAAC GCTATATCCC
 27301 GGGTCAACGG AATCCGCGCC CACCGAAAAC GAATTTCTCT GGAACAGCGC GCTATTACCA
 27361 CCACACCTCG TAATAACCTT AATCCCGTGA GTTGGCCCGC TGCCCTGGTG TACGACGAAA
 27421 GTCCCGCTCC CACCACTGTG GTACTTCCCA GAGACAGCGA GGCCCGAAGT CAGGATGACTA
 27481 ACTCAGGGGC CAGCTTGGG CGCGGCTTTC GTCAAGGGT CGCGTCCGCC GGGCAGGGTA
 27541 TAACCTACCT GACAAATCAGA GGGCGAGGTA TCTCAGCTCA CAGCAGGTGT GTGAGCTTCT
 27601 CGCTTGGTCT CCGTCGGGAC GGGACATTTC AGATCGCGCG CGCGCGCGCT CTTGATACCA
 27661 CGCCTCTGTA GGCATTCCTA ACTCTGCAGA CCTCGTCTCT TGAGCCGCGC TGTGAGGGA
 27721 TTGGAACTCT CAAATTTAT TGGAGGTTTG TGCCATCGCT CTACTTTAAC CCTTCTTCGG
 27781 GACCTCCCGC CCACTATCCG GATCAATTTA TTCTTAACCT TGACGCGGTA AAGGACTTCG
 27841 CGAGCGGCTA CGACTGAATG TTAAGTGGAG AGGCAGAGCA ACTGCGCTG AAACACTCTT
 27901 TCCACTGTGC CGCGCACAA TGCCTTGGCC CGGACTCCGG TGAGTTTTCG TACTTTGAAT
 27961 TGCCCGAGGA TCATATCGAG GCGCCGCGCC GCGCGCTCCG GCTTACCGCC CAGGGAGAGC
 28021 TTGCCCTAG CCGTATTCGG GAGTTTACCC AGCGCCCGCT GCTAGTTTGA CGGGACAGGG
 28081 GACCTCTGT TCTCACTGTG ATTTGCAACT GTCTTAACCT TGGATTAACAT CAAGACTCTT
 28141 GTTGCCATCT CTGTGCTGAG TATAATAAAT ACAGAAATTA AAATATACTG GGGCTCTAT
 28201 CGGCATCTGT TAACGCCAC CCGTCTCAAC CGCCCAAGCA AACCAAGGCG AACCTTACCT
 28261 GGTACTTTTA ACATCTCTCC CTCTGTGATT TACAACAGTT TCAACCCAGA CCGAGTGAGT
 28321 CTACGAGAGA ACCTCTCCGA GCTCAGCTAC TCCATCAGAA AAAACACCA CCGCTTACC
 28381 TGCCGGGAGC GTACGAGTGC GTCAACCGCC GCTGCACCA ACCTACCGCC TGAACGTAAG
 28441 CCAGACTTTT TCCGGACAGA CCTCAATAAC TCTGTTTACC AGAACAGGAG GTGAGCTTAA
 28501 AAAACCTTTA GGGTATTAGG CCAAAGGCGC AGCTACTGTG GGGTTTATGA ACAATTCAG
 28561 CAACCTCTAG GGCATTTCTA ATTCAAGTTT CTCTAGAATC GGGGTGTGGG TTATTTCTTG
 28621 TCTTTGATTT CTCTTTATTC TTATACTAAC GCTTCTCTGC CTAAGGCTCG CGGCTCGTG
 28681 AGGTGACATT TGCATTTATT GTCACTTTT TAAACGCTGG GGTGCGCAC CAAGATGATT
 28741 AGGTACATAA TCTTAGGTTT ACTCACCTTT CGGTGAGCCC ACGTACCAC CCAAAAGGTT
 28801 GATTTTAAAG AGCCAGCCTG TAATGTACCA TCTGCACTG AAGCTAATGA GTGCACACT
 28861 CTATATAAAT GCACCACAGA ACATGAAAG CTGCTTATTC GCCACAAAA CAAATTTGCG
 28921 GATTGTGCTG TTATGCTAT TTTGGCAGCA GTGCACACTA CAGAGTATAA TOTTTACAGT
 28981 TTCCAGGGTA AAGTCAATAA AACTTTTATG TATACTTTTC CATTTTATGA AATGTGCGAC
 29041 ATTACCATGT ACATGAGCAA ACGATATAAG TTTGGGCCCC CACAAAAATG TGTGAAAAAC
 29101 ACTGCGACTT TCTGTGCAAC TGCTATGCTA ATTACAGTGC TCGCTTTGTT CTTGTACCTA
 29161 CTCTATATTA AATACAAAA GAGAGCGAGC TTTATTGAGG AAAAAGAAAT CTTTAAATTT
 29221 ACTTAGTTAC AAGGCTAATG TCACCACATA CTGCTTTACT CGCTGCTTGC AAAACAAAT
 29281 CAAAAAGTTA GCATTTAAT TAAATAGGA TTTAAACCCC CGGTCATT TTCTGCTCAAT
 29341 ACCATTTCCC TGAACAATTG ACTCTAGTTG GGAATATGCT CAGCGCTACA ACCTTGAAT
 29401 CAGGCTCTCT GGATGTCAAG ATCTGACTTT GGCCAGCAC TGTCGCCGAG ATTGTGTTCA
 29461 GTCCACACTAC AGCGACCCAC CCTAACAGAG ATGACCAACA CAACCAACCG GCGCGCGCT
 29521 ACGGACTCTA CATCTACCAC AANTACACC CAAGTTTCTG CTTTGTCAA TAACTGGAT
 29581 AACTTGGGCA TGTGTTGGTT CTCCATAGCG CTTATGTTTG TATGCTTAT TATTATGCG
 29641 CTCACTGCTC GCTTAAAGCG CAAAGCGGCC CGACCACCCA TCTATAGTCC CATCATPTGG

FIG. 81

66/92

29701 CTACACCCAA ACAATGATGG AATCCATAGA TTGGACGGAC TGAAACACAT GTTCTTTTCT
 29761 CTACAGTAT GATTAAATGA GACATGATTG CTCGAGTTT TATATTACTG ACCCTGTGTG
 29821 CGCTTTTTTG TGGCTGTCTC ACATTGGCTG CGGTTTCTCA CATCGAAGTA GACTGCAATC
 29881 CAGCCTTCAC AGTCTATTTG CTPTACGGAT TCTTACCCCT CACGCTCATC TCGAGCCTCA
 29941 TCACGTGGGT CATCGCCTTT ATCCAGTGCA TTGACTGGGT CTGTGTGCGC TTGTGCATATC
 30001 TGACACACA TCCCCAGTAC AGGGACAGGA CTATAGCTGA GCTTCTTAGA ATTCTTTAAT
 30061 TATGAAATTT ACTGTGACTT TTCTGCTGAT TATTTGCACC CTATCTGCGT TTTGTTCCCC
 30121 GACCTCCAAAG CCTCAAGAC ATATATCATG CAGATTCACT CGTATATGGA ATATTCCAA
 30181 TTGCTACAAAT GAAAAAGCG ATCTTTCCGA AGCCTGGTTA TATGCAATCA TCTCTGTATT
 30241 GGTGTTCTGC AGTACCATCT TAGCCCTAGC TATATATCCC TACCTTGACA TTGGCTGGAA
 30301 ACGAATAGAT GCCATGAACC ACCCAACTTT CCGCGGCCCC GCTATGCTTC CACTGCAACA
 30361 AGTTGTGTC GCGGCGTTTG TCCAGCCAA TCAGCCTGCG CCCACTTCTC CACCCCCAAC
 30421 TGAATCAGC TACTTTAATC TAACAGGAGG AGATGACTGA CACCCTAGAT CTAGAAATGG
 30481 ACGGAATTAT TACAGAGCAG CGCCTGCTAG AAAGACGCG GGCAGCGGCC GAGCAACAGC
 30541 GCATGAATCA AGAGCTCCAA GACATGGTTA ACTTGACCA GTGCAAAAAG GGTATCTTTT
 30601 GTCTGGTAAA GCAGGCCAAA GTCACTTACG ACAGTAATAC CACCGGACAC CGCTTAGCT
 30661 ACAAGTTGCC AACCAGCGT CAGAAATGG TGGTCACTGG GGGAGAAAAG CCCATTACCA
 30721 TAACTGACCA CTGCGTAGAA ACCGAGGCT GCACTACTC ACCTTGTCAA GGACTCTGAGG
 30781 ATCTCTGCAC CCTTATTAAG ACCCTGTGCG GTCTCAAAGA TCTTATTCCC TTTTACTTAAT
 30841 AAAAAAATAA ATATAAGCAT CACTTACTTA GATCACTGTA GCAAAATTTCT GTCCAGTTAT
 30901 TTGACAGCA CCTCTTGCC CTCTCCGAG CTCTGGTATT GTTCTGTCTT CGAGCTTCCCT CTGGCTGCA
 30961 AACCTTCTCC ACAATCTAAA TGGAAATGTA GTTCTCTCT GTTCTGTCTT ATCCGCAACC
 31021 ACTATCTTCA TGTGTTGCA GATGAAGCGT GCAAGACCGT CTGAAGATAC CTCTCAACCC
 31081 GTCTGCTTCA ATGACACGGA AACCGTCTCC CTCACTGTGC CTPTTCTTAC TTCTCCCTTT
 31141 GTATCCCCCA ATGGTTTCA AGAGAGTCC CTTGGGTTAC TCTCTTTGGC CTTATCCGAA
 31201 CCTCTAGTTA CCTCCAATGC CATGCTTGGC CTCAAAATGG GCAACGGCC CTCTCTGGAC
 31261 GAGGCCGGCA ACCCTAACCT CAAAAATGTA ACCACTGTGA GCCCACTCT CTAAAAAAC
 31321 AAGTCAAAAC TAAACTTGA AATATCTGCA CCCCTCACAG TACCTCTGCA AGCCCTACT
 31381 GTGGCTGCCG CCGCACTCT ATGGTGCGG GGCACACAG TCACCTAGTA ATCAAGGCC
 31441 CCGCTAACCG TGCACGACTC CAACTTAGC ATTGCCACCC AAGGACCCCT CACAGTGTCA
 31501 GAAGGAAAGC TAGCCTTGCA AACATCAGGC CCCCTCACCA CCACGATAG CAGTACCTT
 31561 ACTATCACTG CCTCACCCCT TCTAACTACT GCCACTGGTA GCTTGGGCA TGTACTGAAA
 31621 GAGCCCATTT ATACACAAAA TGGAAAACTA GGACTAAAT ACGGGGCTCC TTGTGATGTA
 31681 ACAGACGACC TAAACACTTT GACCGTAGCA ACTGGTCCAG GTGTGACTAT TAATAATACT
 31741 TCCCTGCAAA CTAAAGTTAC TGGAGCCTTG GGTTTTGATT CACAAGSCAA TATGCAACTT
 31801 AATGTAGCAG GAGGACTAAG GATTGATTCT CAAAACAGAC GCCTTATATCT TGATGTTGAT
 31861 TATCCGTTTG ATGCTCAAAA CCAACTAAAT CTAAGACTAG GACAGGGCCC TCTTTTATTA
 31921 AACTCAGCCC ACAACTTGA TATTAACTAC AACAAAGGCC TTTACTTTGT TACAGTCTGT
 31981 AACAAATCCA ABAAGCTTGA GGTAACTTA AGCATGCCA AGGGGTTGAT GTTGTAGCTC
 32041 ACAGCCATAG CCAATTATGC AGGAGATGGG CTGAAATTTG GTTCACTTAA TGCACCAAC
 32101 ACRAATCCCC TCAAAACAAA AATGGGCCAT GGCTCTAAGT TTGATTCAAA CCAAGCTATG
 32161 GTTCTTAAAC TAGGAACCTG CCTTAGTTT GACAGCACAG GTGCCATTAC AGTAGGAACA
 32221 AAAATAATG ATAACTTAAC TTTTGGGACC ACACAGCTC CATCTCCATA CTGTAGACTA
 32281 AATGACAGAGA AAGATGTCAA ACTCACTTTG GTCTTAAACA AATGTGGCAG TCAATATCT
 32341 GCTACAGTTT CAGTTTGGC GTTTAAAGGC AGTTTGGCTC CAATATCTGG AACAGTCAA
 32401 AGTGCTCATC TTATTATAAG ATTTGACGAA AATGGAGTGC TACTAAACAA TTCTTCTCTG
 32461 GACCCCAAAAT ATTTGGAATT TAGAAATGGA GATCTTACTG AAGGCACAGC CTATACAAAC
 32521 GCTGTGTGAT TTATGCTTAA CCTATCAGCT TATCCAAAA CTCAAGGTAA AACTGCCAAC
 32581 AGTAACTATG TCACTCAAGT TTAATTAAAC GGAGACAAAA CTAACCTGT CTAACATAAC
 32641 ATTAACATAA ACGGTACACA GGAACACGGA GACACAACTC CAAGTGCATAT CTCTATGTCA
 32701 TTTTCAATGG ACTGGCTGCG CCACAACTAC ATTAATGAAA TATTGTCAC ATCTCTCTAC
 32761 ACTTTTTCAT ACATTGCCCA AGAATAAAGA ATCGTTTGTG TTATGTTTCCA ACGTGTATT
 32821 TTTTCAATGG CAGAAAAATT TCAATTCAGT AGTATAGCCC CACCACCAGC
 32881 TAGCTTATAC AGATCACCGT ACCTTAATCA AACTCACAGA ACCCTAGTAT TCAACCTGCC
 32941 ACCTCCCTCC CAACACACAG AGTACACAGT CCTTTCTGCC CCGCTGCGCT TAAAAAGCAT

FIG. 8J

67/92

33001 CATATCATGG GTAACAGACA TATTCTTAGG TGTATATTC CACACGGTTT CCTGTCGAGC
 33061 CAAACGCTCA TCACTGATAT TAATAAACTC CCGGCGCAGC TCACCTTAAGT TCATGTCGCT
 33121 GTCCACGCTGC TGAGCCACAG GCTGCTGTCC AACTTGCCTG TGCCTTAACGG GCGCGGAAGG
 33181 AGAAGTGCAC GCGTACATGG GGTAGAGTGC ATAACTGTGG ATCAGGATAG GCGGCTGGTG
 33241 CTGCAGCAGC GCGCGAATAA ACTGCTGCCG CCGCCGCTCC GTCCCTGCAG AATACAACAT
 33301 GGCAGTGTCT TCTTCAGCGA TGAATTCGCA CGCCCGCAGC ATAAGGCGCCG TTGTCTTCGC
 33361 GCGCAGCAGC CGCACCTCTA TCTCACTTAA ATCAGCAGAG TAACTGCGAG ACAGCACCCAC
 33421 AATATTGTTTC AAAATCCCAAC AGTGCAGGCG GGTGTATCCA AAGCTCATGG CCGGGACACC
 33481 AGAACCCACG TGGCCATCAT ACCACAAGCG CAGGTAGATT AAGTGGCGAC CCCCTATAAA
 33541 CAGGCTGGAC ATAAACATTA CCTCTTTTGG CATGTTGTAA TTCACCACCT CCCGTTACCA
 33601 TATAAACCTC TGATTAAACA TGGCGCCATC CACCACCATC CTAACCCAGT CCGGTCAACC
 33661 CTGCCCCCGG GCTATACACT GCAGGGAACC GGGACTGGAA CAATGACAGT GGAAGACCCA
 33721 GGACTCGTAA CCAATGGATCA TCAATGCTGG CATGATATCA ATGTGTGGAC AACACAGGCA
 33781 CAGGTGCATA CACTTCTCTA GGATTAACAG CTCTCTCCGC GTTAGAAGCA ATACACAGCA
 33841 AACACCCCAT TCTTGAATCA CGGTAATTCG CACACTGCAG GGAAGACCTC GCACGTAACT
 33901 CAGGTTGTGC ATTGTCAAAG TGTACATTC GGGCAGCAGC GGATGATCTT CCAGTATGTT
 33961 AGCGCGGGTT TCTGTCTCAA AAGGAGGTTAG ACAGTCCCTA CTGTACGGAG TGCGCCGAGA
 34021 CAACCGAGAT CGTGTGTGTC GTAGTGTTCAT GCCAATGGA ACGCCGAGAG TAGTCATATT
 34081 TCTTGAAGCA AAACCAGGTG CGGCGGTGAC AACAGATCTC GGTCTCCGCG TCTGCGCTG
 34141 TAGATCGCTC TGTGTAGTAG TTGTAGTATA TCCACTCTCT CAAAGCATCC AGGCCGCCCC
 34201 TGGCTTCGGG TTCTATGTAA ACTCCTTCTC GCGCCGCTGC CTTGATAACA TCCACCACCG
 34261 CAGATAAGC CACACCCAGC CAACCTACAC ATTCTGTTCT CGAGTCAAC ACGGGAGGAG
 34321 CGGGAAGAGC TGGGAAGAAC ATGTTTTTTT TTTTATTCCA AAAGATTATC CAAAACCTCA
 34381 AAATGAAGAT CTATTAGTGT AACCGCTTCC CCTCCGGTGG CGTGGTCAAA CTCCTACAGC
 34441 AAAGAACACA TAATGGCATT TGTAAAGTGT TGCACAATGG CTTCCAAAAG GCACAAAGCC
 34501 CTCAGTCCA AGTGAGCTA AAGGCTAAAC CCTTCAGGGT GAATCTCCTC TATAACATT
 34561 CCAGCACTT CAACCATGCC CAAATAATTC TCATCTCGCC ACCTTCTCAA TATATCTCTA
 34621 AGCAAAATCC GAATATTAA TCCGGCCATT GTAAAAATCT GCTCCAGAGC GCCTCCACCC
 34681 TTCAGCTTCA AGCAGCGAAT CATGATTGCA AAAATTCAGG TTCTCTCAG ACCTGTATAA
 34741 GATTCAAAGC CGGAACATTA ACAAAAATAC CGCGATCCCG TAGGTCCCTT CGCAGGGCCA
 34801 GCTGAACATA ATCGTGCAGG TCTGCACGGA CCAGCGCGCG CACTTCCCGC CCAGGAACCT
 34861 TGACAAAGAA ACCCACACTG ATTATGACAC GCATACTCGG AGCTATGCTA ACCAGCGTAG
 34921 CCCCATGTA AGCTTTGTGT CATGGGCGCG GATATAAAAT GCAAGGTGCT GCTCAAAAAA
 34981 TCAGGCAAGC CCTCGCGCAA AAAAGAAAGC ACATCGTAGT CATGCTCATG CAGATAAAGG
 35041 CAGGTAAGCT CCGGAACACC CACAGAAAAA GACACCATTT TTCTCTCAAA CATGCTGCGC
 35101 GGTTCCTGCA TAAACACAAA ATAAAAATAC AAAAAACAT TTAACATTA GAAGCCTGTC
 35161 TTACAAACAG AAAACCAACC CTATAAGCA TAAGACGGAC TACGGCCATG CCGGCGTAGC
 35221 CGTAAAAAAA CTGGTCACCG TGATTAAAAA GCACCAACCGA CAGCTCCTCG GTCATGTTCG
 35281 GAGTCATTAAT GTAAGACTCG GTAACACATC CAGGTTGATT CATCGGTCAG TGTCAAAAAG
 35341 CGACCGAAAT AGCCCGGGGG AATACATACC CGCAGGCGTA GAGACAATAT TACAGCCCCC
 35401 ATAGGAGGTA TAACAAAAAT AATAAGAGAG AAAACACAT AAACACCTGA AAAACCTCTC
 35461 TGCTTAGGCA AAATAGCACC CTCCCGCTCC AGAACACAT ACAGCGCTTC ACAGCGGACG
 35521 CTTACACGTC AGCCTTACCA GTAAAAAGA AAACCTATTA AAAAAACACC ACTCGACACG
 35581 GCAACAGCTC AATCAGTCAC AGTGTAAAAA AGGGCCAAGT GCAGAGCGAG TATATATAGG
 35641 ACTAAAAAAT GACGTAAACG TTAAAGTCCA CAAAAACAC CCAGAAACCC GCACGGAAC
 35701 CTACGCCAGC AAACGAAAGC CAAAAACCC ACAACTTCTC CAAATCGTCA CTTCCGTTTT
 35761 CCGACGTTAC GTAACTTCCC ATTTTAAAGA AACTACAAAT CCCAACACAT ACAAGTATAT
 35821 CCGCCCTAAA ACTACGTCA CCGCCGCCGT TCCACGCCC CGCGCCACGT CACAACCTCC
 35881 ACCCCCTCAT TATCATATTG GCTTCAATCC AAAATAAGGT ATATTATTGA TGATG

FIG. 8K

68/92

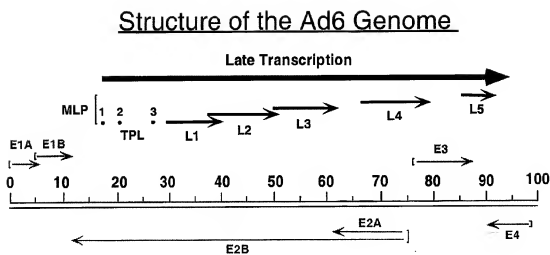


FIG. 9

69/92

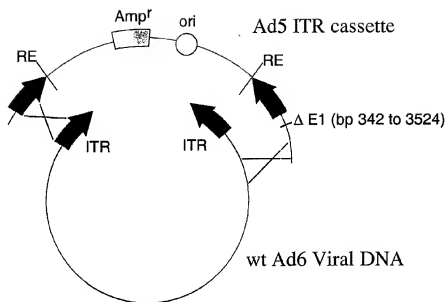


FIG. 10

70/92

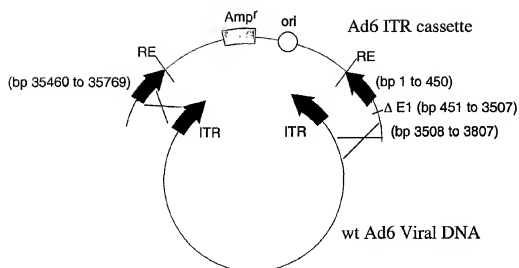
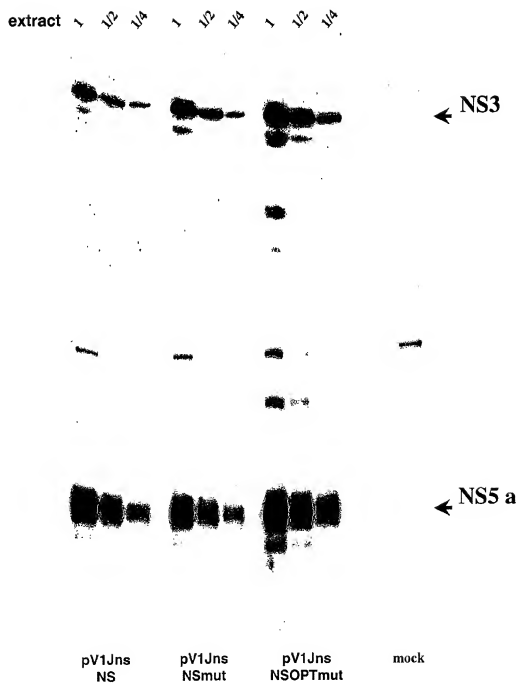


FIG. 11

71/92



Western blot on whole-cell extracts from 293 cells transfected with plasmid DNA expressing the different HCV NS cassettes. Mature NS3 and NS5A products were detected with specific antibodies.

FIG. 12

72/92

		Pep pool						
mouse		F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	1480(CD8 ep)
pV1jns-NS	#31	41	135	19	44	25	17	137
	#32	121	783	77	144	13	22	604
	#33	8	32	3	11	6	6	43
	#34	16	139	13	47	31	25	151
	#35	21	101	40	32	21	20	75
	#36	18	26	24	25	5	7	29
	#37	19	73	15	39	8	20	49
	#38	133	575	74	345	75	63	515
	#39	40	183	10	85	14	9	148
	#40	66	465	29	111	15	16	189
Geomean		33	146	21	57	15	16	123
		Pep pool						
mouse		F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	1480(CD8 ep)
pV1jns-NSmut	#41	39	293	58	187	5	4	248
	#42	21	220	46	107	26	10	189
	#43	76	134	12	78	8	6	144
	#44	30	45	20	52	4	8	40
	#45	36	100	17	56	4	6	116
	#46	67	172	16	138	8	9	145
	#47	34	131	28	38	9	5	118
	#48	55	316	43	107	9	7	277
	#49	8	131	5	25	4	1	91
	#50	13	93	11	11	5	1	78
Geomean		30	142	20	61	7	5	126
		Pep pool						
mouse		F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	1480(CD8 ep)
V1jns-NSOPTmut	#51	53	409	34	84	11	25	271
	#52	140	660	65	276	23	36	377
	#53	58	553	48	105	23	18	564
	#54	50	105	35	134	10	16	80
	#55	14	80	11	35	4	7	91
	#56	14	342	30	101	23	14	207
	#57	63	325	66	239	17	24	123
	#58	75	542	66	168	127	93	191
	#59	65	468	40	124	18	23	344
	#60	27	142	48	16	7	8	77
Geomean		45	295	40	99	16	20	188

IFN γ ELISpot on splenocytes from C57black6 mice immunized with two injections of 25 μ g DNA/dose with GET of plasmid vectors expressing the different HCV NS cassettes. Data are expressed as SFC/10⁶ PBMC.

FIG. 13A

73/92

		Pep pool						
mouse		F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	DMSO
pV1jns-NS	#51	219	699	634	486	487	264	34
	#52	67	302	347	167	111	87	9
	#53	59	460	400	246	244	136	26
	#54	139	817	685	236	547	223	24
	#55	96	904	542	277	256	337	17
	#56	225	603	666	156	350	240	56
	#57	44	288	211	148	100	141	4
	#58	37	262	221	53	58	62	3
	#59	131	975	928	159	305	284	14
	#60	93	475	464	77	206	113	12
geo mean		111	579	512	201	266	189	20

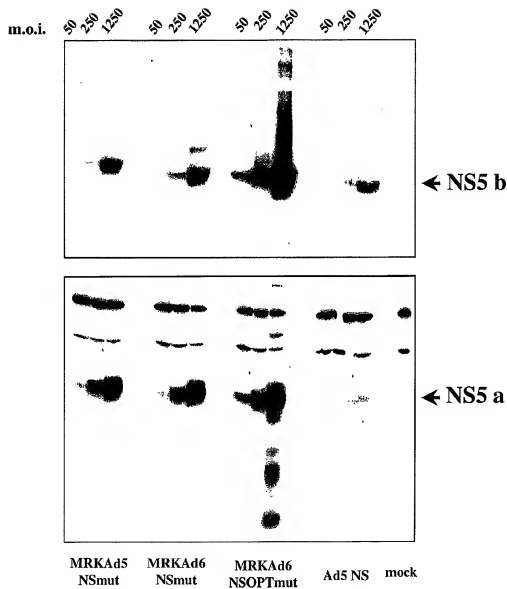
		Pep pool						
mouse		F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	DMSO
pV1jns-NSmut	#61	72	840	515	219	278	249	19
	#62	294	1881	1266	365	434	411	63
	#63	73	415	422	103	141	99	41
	#64	66	824	486	175	162	144	18
	#66	24	313	168	53	47	42	5
	#67	15	230	253	94	25	39	2
	#68	53	354	252	89	101	86	15
	#69	271	895	909	518	322	285	74
	#70	417	1303	1186	468	557	267	34
	geo mean		143	784	606	232	230	180

		Pep pool						
mouse		F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L(NS35b)	M(NS5b)	DMSO
V1jns-NSOPTmut	#71	206	944	890	342	207	397	47
	#72	393	1655	1151	575	626	401	72
	#73	123	522	515	319	223	198	21
	#74	500	1414	1419	878	1035	1122	137
	#75	286	812	873	382	543	267	31
	#76	224	1143	942	218	420	281	22
	#77	95	643	630	169	385	218	15
	#78	401	1302	1068	538	608	623	12
	#79	108	1190	914	199	265	215	4
	#80	122	511	546	189	286	190	13
geo mean		209	941	854	331	406	329	24

IFN γ ELISpot on splenocytes from BalbC mice immunized with two injections of 50 μ g DNA/dose with GET of plasmid vectors expressing the different HCV NS cassettes. Data are expressed as SFC/10⁶ PBMC.

FIG. 13B

74/92



Western blot on whole-cell extracts from HeLa cells infected at different multiplicity of infection (m.o.i.; indicated at the top) with Adenovectors expressing the different HCV NS cassettes. Mature NS5B and NS5A products were detected with specific antibodies.

FIG. 14

75/92

		Pep pool					
		F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L+M(NS35b)	1480(CD8 ep)DMSO
Ad5-NS	mouse						
	#1	14	492	9	27	10	554
	#2	8	440	2	26	5	438
	#3	12	92	5	12	7	73
	#4	16	388	6	40	6	228
	#6	8	210	4	31	3	238
	#7	7	133	13	16	0	128
	#8	11	342	25	55	22	267
	#9	5	345	0	45	5	285
	#10	22	888	3	65	25	799
Geomean		10	305	na	31	na	269
		Pep pool					
		F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L+M(NS35b)	1480(CD8 ep)DMSO
MRKAd5-NSmut	mouse						
	#11	14	1009	13	75	7	751
	#12	15	695	3	39	9	552
	#13	12	389	4	20	7	352
	#14	7	459	6	50	1	274
	#15	5	549	3	22	6	485
	#16	10	631	1	6	4	600
	#17	5	257	3	9	1	245
	#18	13	659	6	43	7	555
	#19	12	758	1	37	5	669
	#20	22	1380	5	163	8	1003
Geomean		10	615	3	31	4	504
		Pep pool					
		F(NS3p)	G(NS3h)	H(NS4)	I(NS5a)	L+M(NS35b)	1480(CD8 ep)DMSO
MRKAd6-NSmut	mouse						
	#21	6	584	5	27	4	491
	#22	6	231	3	12	3	235
	#23	8	482	1	18	1	511
	#24	14	1120	6	38	10	1004
	#25	1	311	3	9	0	382
	#26	29	903	3	60	5	751
	#27	35	1573	4	40	4	1277
	#28	7	406	5	15	1	443
	#29	4	461	3	12	3	515
Geomean		8	567	3	21	na	554

IFN γ ELISPOT on splenocytes from C57black6 mice immunized with two injections of 10^6 vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

FIG. 15

76/92

Pep pools	Ad5-NS 10^{10} vp/dose		
	96074	134T	063Q
<i>F (NS3p)</i>	374	11	74
<i>G (NS3h)</i>	359	1070	1455
<i>H (NS4)</i>	376	30	64
<i>I (NS5a)</i>	240	40	63
<i>L (NS5b)</i>	226	29	121
<i>M (NS5b)</i>	511	23	35
<i>DMSO</i>	128	3	31

Pep pools	MRK Ad6-NSmut 10^{10} vp/dose		
	S207	035Q	057Q
<i>F (NS3p)</i>	363	382	150
<i>G (NS3h)</i>	180	316	119
<i>H (NS4)</i>	126	113	62
<i>I (NS5a)</i>	1780	688	114
<i>L (NS5b)</i>	447	111	81
<i>M (NS5b)</i>	153	38	16
<i>DMSO</i>	9	6	9

IFN γ ELISPOT on PBMC from Rhesus monkeys immunized with one injection of 10^{10} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

FIG. 16A

77/92

Pep pools	MRK Ad5-NSmut 10^{10} vp/dose		
	<i>S201</i>	<i>075Q</i>	<i>I37Q</i>
<i>F (NS3p)</i>	928	69	254
<i>G (NS3h)</i>	317	436	98
<i>H (NS4)</i>	56	101	45
<i>I (NS5a)</i>	1530	1100	413
<i>L (NS5b)</i>	149	23	92
<i>M (NS5b)</i>	398	32	80
<i>DMSO</i>	29	6	29

Pep pools	MRK Ad6-NSOPTmut 10^{10} vp/dose		
	<i>98D209</i>	<i>106Q</i>	<i>I13Q</i>
<i>F (NS3p)</i>	3110	263	404
<i>G (NS3h)</i>	2115	642	1008
<i>H (NS4)</i>	373	72	19
<i>I (NS5a)</i>	103	37	347
<i>L (NS5b)</i>	149	22	10
<i>M (NS5b)</i>	314	428	19
<i>DMSO</i>	0	1	3

IFN γ ELISPOT on PBMC from Rhesus monkeys immunized with one injection of 10^{10} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

FIG. 16B

78/92

Pep pools	Ad5-NS 10^{11} vp/dose			
	99C008	97N104	97X008	99C026
<i>F (NS3p)</i>	28	1026	579	889
<i>G (NS3h)</i>	1279	188	103	2453
<i>H (NS4)</i>	18	39	138	109
<i>I (NS5a)</i>	131	1068	172	141
<i>L (NS5b)</i>	78	144	103	32
<i>M (NS5b)</i>	24	68	47	84
<i>DMSO</i>	3	16	1	19

Pep pools	MRKAd6-NSmut 10^{11} vp/dose			
	98C047	97C055	93G	97X014
<i>F (NS3p)</i>	477	25	93	1022
<i>G (NS3h)</i>	959	398	81	1513
<i>H (NS4)</i>	36	14	99	53
<i>I (NS5a)</i>	171	45	1237	98
<i>L (NS5b)</i>	18	32	23	51
<i>M (NS5b)</i>	88	4	13	40
<i>DMSO</i>	8	3	1	5

IFN γ ELISPOT on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

FIG. 16C

79/92

Pep pools	MRKAd5-NSmut 10^{11} vp/dose			
	99C059	99C060	97K009	96069
<i>F (NS3p)</i>	28	81	1308	1618
<i>G (NS3h)</i>	2600	161	1008	123
<i>H (NS4)</i>	31	74	101	40
<i>I (NS5a)</i>	181	99	69	96
<i>L (NS5b)</i>	24	31	40	20
<i>M (NS5b)</i>	11	58	38	164
<i>DMSO</i>	6	15	1	16

IFN γ ELISPOT on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as SFC/ 10^6 PBMC.

FIG. 16D

80/92

	MRK Ad5-NSmut 10 ¹⁰ vp/dose		
Pcp pools	<i>S201</i>	<i>075Q</i>	<i>137Q</i>
<i>pool F (NS3p)</i>	881	1755	73
<i>pool G (NS3h)</i>	573		.
<i>pool H (NS4)</i>		3541	
<i>pool I (NS5a)</i>	2094		39
<i>pool L (NS5b)</i>			
<i>pool M (NS5b)</i>	756		
<i>DMSO</i>	319	117	44

	MRK Ad6-NSOPTmut 10 ¹⁰ vp/dose		
Pcp pools	<i>98D209</i>	<i>106Q</i>	<i>113Q</i>
<i>pool F (NS3p)</i>	5073	84	952
<i>pool G (NS3h)</i>	2376	160	3325
<i>pool H (NS4)</i>	700		
<i>pool I (NS5a)</i>			1106
<i>pool L (NS5b)</i>			
<i>pool M (NS5b)</i>	530	706	
<i>DMSO</i>	43	47	28

	MRK Ad6-NSmut 10 ¹⁰ vp/dose		
Pcp pools	<i>S207</i>	<i>035Q</i>	<i>057Q</i>
<i>pool F (NS3p)</i>	118	480	
<i>pool G (NS3h)</i>		196	
<i>pool H (NS4)</i>			
<i>pool I (NS5a)</i>	3340	933	
<i>pool L (NS5b)</i>	118		
<i>pool M (NS5b)</i>			
<i>DMSO</i>	145	34	

IFN γ ICS on PBMC from Rhesus monkeys immunized with two injections at four weeks interval with 10¹⁰ vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as number of positive IFN γ /CD3/CD8 per 10⁶ lymphocytes.

FIG. 17A

81/92

Pep pools	Ad5-NS 10 ¹¹ vp/dose			
	99C008	97N104	97X008	99C026
<i>F (NS3p)</i>		1703	1136	615
<i>G (NS3h)</i>	3153			2787
<i>H (NS4)</i>				
<i>I (NS5a)</i>		2233		
<i>L (NS5b)</i>				
<i>M (NS5b)</i>				
<i>DMSO</i>	125	98	130	0

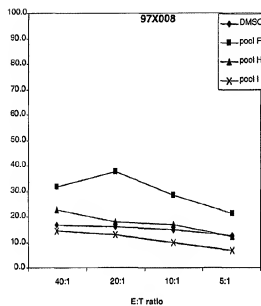
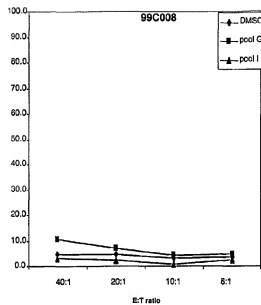
Pep pools	MRKAd6-NSmut 10 ¹¹ vp/dose			
	98C047	97C055	93G	97X014
<i>F (NS3p)</i>	1024			948
<i>G (NS3h)</i>	3246	353		1074
<i>H (NS4)</i>			316	
<i>I (NS5a)</i>			6224	
<i>L (NS5b)</i>				
<i>M (NS5b)</i>				
<i>DMSO</i>	49	23	37	93

Pep pools	MRKAd5-NSmut 10 ¹¹ vp/dose			
	99C059	99C060	97X009	96069
<i>F (NS3p)</i>			2266	5053
<i>G (NS3h)</i>	2434	316	1018	
<i>H (NS4)</i>				
<i>I (NS5a)</i>				
<i>L (NS5b)</i>				
<i>M (NS5b)</i>				205
<i>DMSO</i>	13	110	119	15

IFN γ ICS on PBMC from Rhesus monkeys immunized with two injections at four weeks interval with 10¹¹ vp/dose of Adenovectors expressing the different HCV NS cassettes. Data are expressed as number of positive IFN γ /CD3/CD8 per 10⁶ lymphocytes.

FIG. 17B

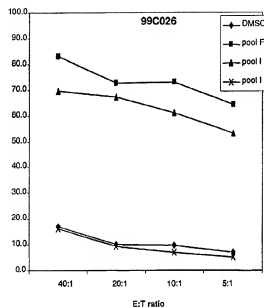
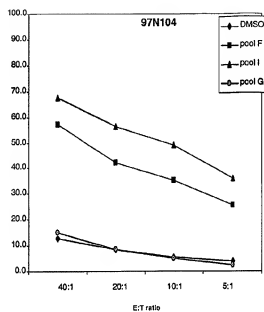
82/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of Ad5-NS.

FIG. 18A

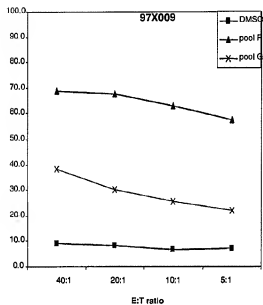
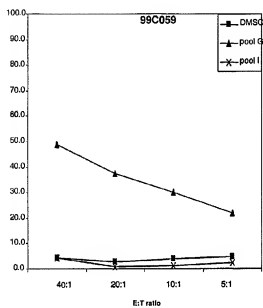
83/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of Ad5-NS.

FIG. 18B

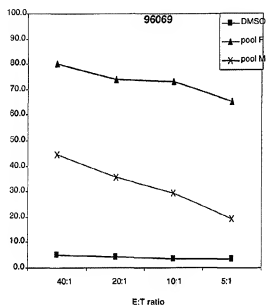
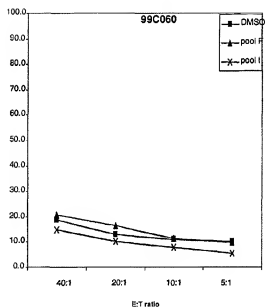
84/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of MRKAd5-NSmut.

FIG. 18C

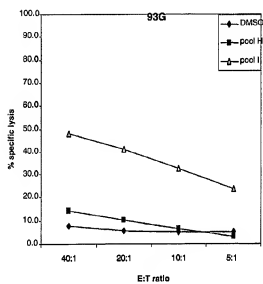
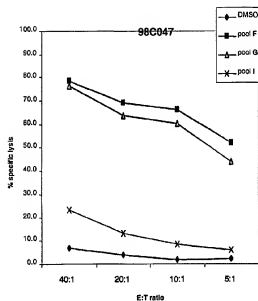
85/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of MRKAd5-NSmut

FIG. 18D

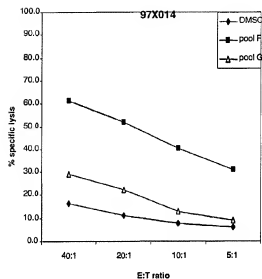
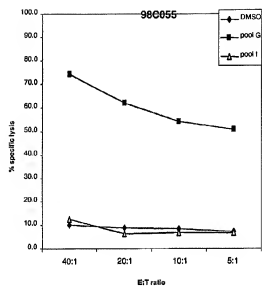
86/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of MRKAd6-NSmut.

FIG. 18E

87/92



Bulk CTL assays on PBMC from Rhesus monkeys immunized with two injections of 10^{11} vp/dose of MRKAd6-NSmut.

FIG. 18F

88/92

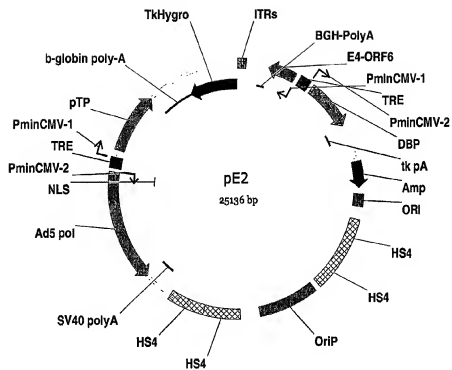


FIG. 19

89/92

1 GCCACCATGG CCCCCATCAC CGCTACAGC CAGCAGACCA GGGGCTGTCT
 51 GGGCTGCATC ATCACCAGCC TGACCGGACG CGACAAGAAC CAGGTGGAGG
 101 GAGAGGTGCA GGTGGTGAGC ACCGCTACCC AGAGCTTCCT GCCCACCTGC
 151 GTGAACGGCG TGTGCTGGAC CGTGATACCAC GGAGCCGGAA GCAGAACCCT
 201 GGC CGGACCC AAGGCCCCA TCACCCAGAT GTACACCAAT GTGGATCAGG
 251 ATCTGGTGGG CTGGCAGGCC CCTCCGGAG CCAGGAGCCT GACACCCCTGT
 301 ACCTGTGGAA GCAGCGACCT GTACCTGGTG ACACGCCACG CCGATGTGAT
 351 CCCCCTGAGG CGCAGGGGCG ATTCTCGCGG AAGCCTGCTG AGCCCTAGGC
 401 CCGTGAGCTA CCTGAAGGGC AGCAGCGGAG GACCCCTGCT GTGTCTTCT
 451 GGC CATGCCG TGGGCATTTT TCGCCTGCC GTGTGTACCA GGGCGGTGGC
 501 CAAAGCGCTG GATTTTGTGC CCGTGGAAG CATGGAGACC ACCATGCGCA
 551 GCCCTGTGTT CACCGACAAC AGCTCTCCCC CTGCCGTGCC CCAATCATTC
 601 CAGGTGGCTC ACCTGCACGC CCTACCCGGA TCTGGCAAGA GCACCAAGGT
 651 GCCCCTGCTC TACGCCGCTC AGGCTACAA GGTGCTGGTG CTGAACCCCA
 701 GCTGGCCGC TACCCTGGGC TTCGGCGCTT ACATGAGCAA GGCCCATGGC
 751 ATCGACCCCA ACATCCGAC AGGCGTGCGC ACCATCACA CCGGAGCTCC
 801 CGTGACCTAC AGCACCTACG GCAAGTTCCT GGCCGATGGA GGCTGCAGCG
 851 GAGGAGCCTA CGACATCATC ATCTGCGACG AGTGCCACAG CACCGACAGC
 901 ACCACCATCC TGGGCATGG CACCGTGCTG GATCAGGCCG AAACAGCTGG
 951 AGCCAGGCTG GTGGTGCTGG CCACAGCTAC CCTCTCTGGC AGCGTGACCG
 1001 TGCCCCATCC CAATATCGAG GAGGTGGCCC TGAGCAACAC AGGCGAGATC
 1051 CCTTCTTACG GCAAGGCCAT CCCCATCGAG GCCATCCGCG GAGGACGGCA
 1101 CCTGATCTTC TGCCACAGCA AGAAGAAAGTG CGACGAGCTG GCTGCCAAGC
 1151 TGAGCGGACT GGGCATCAAC GCCGTGGCCT ACTACAGGGG CTGGACGTG
 1201 TCAGTGATCC CCACCATCGG CGATGTGGTG GTGGTGGCCA CCGACGCCCT
 1251 GATGACAGGC TACACCGGAG ACTTCGACAG CGTGATCGAC TGCAACACCT
 1301 GCGTGACCCA GACCGTGGAC TTCAGCCTGG ACCCCACCTT CACCATCGAA
 1351 ACCACCACCG TGCCCTCAGGA TGCTGTGAGC AGGAGCCAGA GGC GCGGACG
 1401 CACCGGAAGG GGCAGGCGCG GAATTTATCG CTTTGTGACC CTGGCGAAA
 1451 GGCCCTCTGG CATGTTGAC AGCAGCGTGC TGTGCGAGTG CTACGACGCT
 1501 GGC TGCGCTT GGTACGAGCT GACACCCGCT GAAACAGCG TGGCCTTGG
 1551 CGCTTATCTG AATACCCCTG GCCTGCCCGT GTGTCAGGAC CACCTGGAGT

FIG. 20A

90/92

1601 TCTGGGAGAG CGTGTTCACA GGACTGACCC ACATCGAGC CCATTTCCTG
 1651 AGCCAGACCA AGCAGGCTGG CGACAACCTC CCCTATCTGG TGGCCTATCA
 1701 GGCCACCCTG TGTGCTAGGG CCCAAGCTCC ACCTCCTTCA TGGGACCAGA
 1751 TGTGGGAAGTG CCTGATCCGC CTGGAAGCCA CCTGCACGG CCTTACCCTT
 1801 CTGCTGTACC GCCTGGGAGC CGTGCAGAAC GAGGTGACCC TGACCCACCC
 1851 CATACCAAG TACATCATGG CCTGCATGAG CGCTGATCTG GAAGTGGTGA
 1901 CCAGCACCTG GGTGCTGGTG GGAGGCGTGC TGGCGCTCT GGCTGCCTAC
 1951 TGCCTGACCA CCGGAAGCGT GGTGATCGTG GGAGCATCA TCCTGAGCGG
 2001 AAGGCCCGCT ATCGTGCCCG ATGCGGAGTT CCTGTACCAG GAGTTCGACG
 2051 AGATGGAGGA GTGTGCCAGC CACCTGCCCT ACATCGAGCA GGGCATGCAG
 2101 CTGGCCGAAC AGTTCAAGCA GAAGGCCCTG GGCCTGCTGC AGACAGCCAC
 2151 CAAACAGGCC GAAGCTGCCG CTCCCCTGGT GGAAGCAAG TGGAGGGCCC
 2201 TGGAGACCTT CTGGGCTAAG CACATGTGGA ACTTCATCTC TGGCATCCAG
 2251 TACCTGGCCG GACTGAGCAC CTGCTCTGGC AACCCTGCTA TCGCCAGCCT
 2301 GATGGCCTTC ACCGCTAGCA TCACCTCTCC CCTGACCACC CAGAGCACCC
 2351 TGCTGTTCAA CATTCTGGGC GGATGGGTGG CGCTCAGCT GGCCCTCTCT
 2401 TCAGCTGCTT CTGCCCTTGT GGGCGCTGGC ATTGCCGGAG CCGCTGTGGG
 2451 CAGCATTGGC CTGGGCAAAG TGCTGGTGA TATTCTGGCT GGCTATGGCG
 2501 CTGGCGTGGC CGGAGCCCTG GTGGCCTTCA AGGTGATGAG CGGAGAGATG
 2551 CCCAGCACCG AGGACCTGGT GAACCTGCTG CCTGCCATTC TGAGCCCTGG
 2601 AGCCCTGGTG GTGGCGGTGG TGTGTGCTGC CATTCTGAGG CGCCTATGTG
 2651 GACCCGGAGA GGGCGCTGTG CAGTGGATGA ACCGCTGAT CGCCTTCGCC
 2701 TCTCGCGGAA ACCACGTGAG CCTTACCCAC TACGTGCTTG AGAGCGAGCG
 2751 CGCTGCCAGG GTGACCCAGA TCTGAGCAG CCTGACCATC ACCCAGCTGC
 2801 TGAAGCGCCT GCACCACTGG ATCAACGAGG ACTGCAGCAC ACCCTGCAGC
 2851 GGAAGCTGGC TGAGGGACGT GTGGGACTGG ATCTGCACCG TGCTGACCGA
 2901 CTTCAAGACC TGGCTGCAGA GCAAGCTGCT GCCCAACTG CCTGGCGTGC
 2951 CCTTCTTCTC ATGCCAGCGC GGATACAAGG GCGTGTGGAG GGGCGATGGC
 3001 ATCATGCAGA CCACCTGTCC CTGCGGAGCC CAGATCACAG GCCACGTGAA
 3051 GAACGGCAGC ATGCGCATCG TGGGCCCTAA GACCTGCAGC AACACCTGGC
 3101 ACGGCACCTT CCCCATCAAC GCCTACACCA CCGGACCCCTG CACACCCAGC
 3151 CCTGCTCCCA ACTACAGCAG GGCCCTGTGG AGGGTGGCTG CCGAGGAGTA

FIG. 20B

91/92

3201 CGTGGAGGTG ACCAGGGTGG GAGACTTCCA CTACGTGACC GGAATGACCA
 3251 CCGACAACGT GAAAGTGCCC TGTACAGTGC CCGCTCCCGA ATTTTTTACC
 3301 GAAAGTGGATG GCGTGGCGCT GCATCGCTAT GCCCTGCGCT GTAGGCCCCCT
 3351 GCTGCGCGAA GAAAGTGACCT TCAGGTGGG CTTGAACCAG TACCTGGTGG
 3401 GCAGCCAGCT GCCCTGCGAG CCTGAGCCCG ATGTGGCGT GCTGACCAGC
 3451 ATGCTGACCG ACCCCAGCCA CATCACAGCC GAAACCGCTA AAAGCGCGCT
 3501 GGGCAGGGGC TCTCTCCAA GCCTGGCCTC AAGCAGCGCT AGCCAGCTGT
 3551 CTGCTCCAG CCTGAAGGCC ACCTGCACCA CCCACCACGT GAGCCCCGAC
 3601 GCCGACCTGA TCGAGGCCAA CCTGCTGTGG CGCCAGGAGA TGGCGGCAA
 3651 CATCACCCGC GTGGAGAGCG AGAACAAAGT GGTGGTGCTG GACAGCTTCG
 3701 ACCCCCTGCG CGCGAGGGAG GACGAGCGCG AGGTGAGCGT GCCCGCCGAG
 3751 ATCTGCGCA AGAGCAAGAA GTTCCCCGCT GCCATGCCCC TCTGGGCTAG
 3801 ACCTGATTAC AACCCCTCCC TGCTGGAGAG CTGGAAGGAC CCTGATTACG
 3851 TGCTCCAGT GGTGCATGGC TGTCTCTGCG CTCCCATTAA AGCCCTCTCT
 3901 ATTCCACCTC TAGGCGCAA AAGGACCGTG CTGCTGACAG AAAGCAGCGT
 3951 GAGCTCTGCT CTGGCCGAAC TGGCCACCAA GACCTTTGCG AGCAGCGAGA
 4001 GCTCTGCCGT GGACAGCGGA ACAGCCACCG CTCTGCCTGA CCAGGCCAGC
 4051 GACGACGGCG ATAAGGGCAG CGATGTGGAG AGCTATAGCA GCATGCCTCC
 4101 CCTGGAAGGC GAACCTGGCG ATCCCGATCT GAGCGATGGC AGCTGAGCA
 4151 CCGTGAGCGA AGAGGCCAGC GAGGACGTGG TGTGTTGCAG CATGAGCTAC
 4201 ACCTGGACAG GCGCTCTGAT CACACCTTGC GTTGCCGAGG AGAGCAAGCT
 4251 GCCCATCAAC GCCCTGAGCA ACAGCCTGCT GAGGCACCAC AACATGTTGT
 4301 ACGCCACCAC CAGCAGGTCT GCCGGACTGA GGCAGAAGAA GGTGACCTTC
 4351 GACCGCCTGC AGGTGCTGGA CGACCACTAC CGCGATGTGC TGAAGGAGAT
 4401 GAAAGCCAG GCCAGCACCG TGAAGGCCAA GCTGCTGAGC GTGGAGGAGG
 4451 CCTGCAAGCT GACCCCCCCC CACAGCGCCA AGAGCAAGTT CGGCTACGGC
 4501 GCCAAGGACG TGGCAACCT GAGCAGCAAG GCCGTGAACC ACATCCACAG
 4551 CGTGTGGAAG GACCTGCTGG AGGACACCGT GACCCCATC GACACCACCA
 4601 TCATGCCCCA GAACGAGGTG TTCTGCGTGC AGCCCGAGAA GGGCGCCCGC
 4651 AAGCCCGCTC GCCTGATCGT GTTCCCGAT CTGGCGGTGC GCCTGTGCGA
 4701 GAAGATGGCC CTGTACGACG TGGTGAGCAC CCTGCCTCAG GTGGTGATGG
 4751 GCTCAAGCTA CGGCTTCCAG TACAGCCCTG GCCAGCGCGT GGAGTTCCTG

FIG. 20C

92/92

4801 GTGAACACCT GGAAGAGCAA GAAGAACCC ATGGGCTTCA GCTACGACAC
 4851 ACGCTGCTTC GACAGCACC GACCCGAGAA CGACATCCGC GTGGAGGAGA
 4901 GCATCTACCA GTGCTGCGAC CTGGCCCCTG AGGCCAGGCA GGCCATCAAG
 4951 AGCCTGACCG AGCGCTGTA CATCGGAGGC CCTCTGACCA ACAGCAAGGG
 5001 ACAGAACTGC GGATACAGGC GCTGTAGGGC CTCTGGCGTG CTGACCACCA
 5051 GCTGTGGCAA CACCTGACC TGCTACCTGA AGGCCAGCGC TGCTGTCCG
 5101 GCTGCCAAGC TGCAGGACTG CACCATGCTG GTGAACGCCG CTGGCTGTGT
 5151 GGTGATTTGT GAAAGCGCTG GCACCCAGGA AGATGCTGCC AGCCTGCGCG
 5201 TGTTCACCGA GGCCATGACC AGGTACTCTG CCCCTCCCGG AGACCCCCCT
 5251 CAGCCCGAAT ACGACCTGGA GCTGATCACC AGCTGCTCAA GCAACGTGAG
 5301 CGTGGCTCAC GACGCCAGCG GAAAGCGCGT GTACTACCTG ACACGCGATC
 5351 CCACCACCCC TCTGGCTCGC GCTGCCCTGG AAACCGCTCG CCATACACCC
 5401 GTGAACAGCT GGCCTGGCAA CATCATCATG TACGCCCTTA CCCTGTGGGC
 5451 TCGCATGATC CTGATGACCC ACTTCTTCAG CATCCTGCTG GCTCAGSAGC
 5501 AGCTGGAGAA GGCCCTGGAC TGCCAGATT ACGGCGCTTG CTACAGCATC
 5551 GAGCCCTGG ACCTGCCCA AATCATCGAG CGCCTGCACG GCCTGTCTGC
 5601 CTTAGCCTG CACAGCTACA GCCCTGGCGA AATTAATCGC GTGGCCAGCT
 5651 GTCTGCGCAA ACTGGGCGTG CCTCCTCTGC GCGTGTGGAG GCATAGGGCT
 5701 AGGAGCGTGA GGGCTAGGCT GCTGAGCCAG GGAGGCAGGG CCGCTACCTG
 5751 TGGAAAGTAC CTGTTCAACT GGGCCGTGAA GACCAAGCTG AAGCTGACCC
 5801 CTATCCCTGC CGCTAGCCAG CTGGACCTGA GCGGATGGTT CGTGGCTGGC
 5851 TACAGCGGAG GCGACATCTA CCACAGCCTG TCTCGCGCTC GCCCTCGCTG
 5901 GTTCATGCTG TGCTGCTGTC TGCTGAGCGT GGGCGTGGGC ATCTACCTGC
 5951 TGCCCAACCG CTA

FIG. 20D

IN THE PCT RECEIVING OFFICE
OF THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s):	Merck & Co., Inc		
PCT Serial No.:	To Be Assigned	Case No.: PCT ITR0015Y	US/RO
Filing date:	On Even Date Herewith		
For:	HEPATITIS C VIRUS VACCINE		
			Authorized Officer: To Be Assigned

Assistant Commissioner of Patents
BOX PCT
Washington, D.C. 20231


**NUCLEOTIDE AND/OR AMINO ACID
SEQUENCE DISCLOSURE, PCT RULE 5.2**

Sir:

As required under PCT Rule 5.2, Applicant respectfully encloses a paper (64 pages) and a computer readable form of the Sequence Listing for the above-identified PCT International Application, filed on even date herewith.

I hereby state that the content of the paper and computer readable forms of the Sequence Listing, submitted in accordance with WIPO and Standard ST.23 and under PCT Rule 13ter.1, respectively, are the same.

Respectfully submitted,

By 
Sheldon O. Heber
Reg. No. 38,179
Attorney for Applicants

Merck & Co., Inc.
P.O. Box 2000
Rahway, NJ 07065-0907
(732) 594-1958

Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr
 210 215 220
 Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly
 225 230 235 240
 Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly
 245 250 255
 Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly
 260 265 270
 Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile
 275 280 285
 Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Thr Ile Leu Gly Ile
 290 295 300
 Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val
 305 310 315 320
 Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn
 325 330 335
 Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly
 340 345 350
 Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe
 355 360 365
 Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly
 370 375 380
 Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val
 385 390 395 400
 Ile Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met
 405 410 415
 Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys
 420 425 430
 Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu
 435 440 445
 Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly
 450 455 460
 Arg Thr Gly Arg Gly Arg Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly
 465 470 475 480
 Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr
 485 490 495
 Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val
 500 505 510
 Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp
 515 520 525
 His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp
 530 535 540
 Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr
 545 550 555 560
 Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro
 565 570 575
 Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr
 580 585 590
 Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn
 595 600 605
 Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met
 610 615 620
 Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly
 625 630 635 640

Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Thr Thr Gly Ser Val Val
 645 650 655
 Ile Val Gly Arg Ile Ile Leu Ser Gly Arg Pro Ala Ile Val Pro Asp
 660 665 670
 Arg Glu Phe Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ala Ser
 675 680 685
 His Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe Lys
 690 695 700
 Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu Ala
 705 710 715 720
 Ala Ala Pro Val Val Glu Ser Lys Trp Arg Ala Leu Glu Thr Phe Trp
 725 730 735
 Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly
 740 745 750
 Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe
 755 760 765
 Thr Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln Ser Thr Leu Leu Phe
 770 775 780
 Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser Ala
 785 790 795 800
 Ala Ser Ala Phe Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly Ser
 805 810 815
 Ile Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala
 820 825 830
 Gly Val Ala Gly Ala Leu Val Ala Phe Lys Val Met Ser Gly Glu Met
 835 840 845
 Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro
 850 855 860
 Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His
 865 870 875 880
 Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala
 885 890 895
 Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu
 900 905 910
 Ser Asp Ala Ala Ala Arg Val Thr Gln Ile Leu Ser Ser Leu Thr Ile
 915 920 925
 Thr Gln Leu Leu Lys Arg Leu His Gln Trp Ile Asn Glu Asp Cys Ser
 930 935 940
 Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile Cys
 945 950 955 960
 Thr Val Leu Thr Asp Phe Lys Thr Trp Leu Gln Ser Lys Leu Leu Pro
 965 970 975
 Gln Leu Pro Gly Val Pro Phe Phe Ser Cys Gln Arg Gly Tyr Lys Gly
 980 985 990
 Val Trp Arg Gly Asp Gly Ile Met Gln Thr Thr Cys Pro Cys Gly Ala
 995 1000 1005
 Gln Ile Thr Gly His Val Lys Asn Gly Ser Met Arg Ile Val Gly Pro
 1010 1015 1020
 Lys Thr Cys Ser Asn Thr Trp His Gly Thr Phe Pro Ile Asn Ala Tyr
 1025 1030 1035 1040
 Thr Thr Gly Pro Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg Ala
 1045 1050 1055
 Leu Trp Arg Val Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val Gly
 1060 1065 1070

Asp Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Val Lys Cys Pro
 1075 1080 1085
 Cys Gln Val Pro Ala Pro Glu Phe Phe Thr Glu Val Asp Gly Val Arg
 1090 1095 1100
 Leu His Arg Tyr Ala Pro Ala Cys Arg Pro Leu Leu Arg Glu Glu Val
 1105 1110 1115 1120
 Thr Phe Gln Val Gly Leu Asn Gln Tyr Leu Val Gly Ser Gln Leu Pro
 1125 1130 1135
 Cys Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp
 1140 1145 1150
 Pro Ser His Ile Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg Gly
 1155 1160 1165
 Ser Pro Pro Ser Leu Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro
 1170 1175 1180
 Ser Leu Lys Ala Thr Cys Thr Thr His His Val Ser Pro Asp Ala Asp
 1185 1190 1195 1200
 Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile
 1205 1210 1215
 Thr Arg Val Glu Ser Glu Asn Lys Val Val Val Leu Asp Ser Phe Asp
 1220 1225 1230
 Pro Leu Arg Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala Glu
 1235 1240 1245
 Ile Leu Arg Lys Ser Lys Lys Phe Pro Ala Ala Met Pro Ile Trp Ala
 1250 1255 1260
 Arg Pro Asp Tyr Asn Pro Pro Leu Leu Glu Ser Trp Lys Asp Pro Asp
 1265 1270 1275 1280
 Tyr Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Ile Lys Ala
 1285 1290 1295
 Pro Pro Ile Pro Pro Pro Arg Arg Lys Arg Thr Val Val Leu Thr Glu
 1300 1305 1310
 Ser Ser Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe Gly
 1315 1320 1325
 Ser Ser Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Leu Pro
 1330 1335 1340
 Asp Gln Ala Ser Asp Asp Gly Asp Lys Gly Ser Asp Val Glu Ser Tyr
 1345 1350 1355 1360
 Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser
 1365 1370 1375
 Asp Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val Val
 1380 1385 1390
 Cys Cys Ser Met Ser Tyr Thr Thr Gly Ala Leu Ile Thr Pro Cys
 1395 1400 1405
 Ala Ala Glu Glu Ser Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu
 1410 1415 1420
 Leu Arg His His Asn Met Val Tyr Ala Thr Thr Ser Arg Ser Ala Gly
 1425 1430 1435 1440
 Leu Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Asp
 1445 1450 1455
 His Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val
 1460 1465 1470
 Lys Ala Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro Pro
 1475 1480 1485
 His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn
 1490 1495 1500

Leu Ser Ser Lys Ala Val Asn His Ile His Ser Val Trp Lys Asp Leu
 1505 1510 1515 1520
 Leu Glu Asp Thr Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn
 1525 1530 1535
 Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg
 1540 1545 1550
 Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala
 1555 1560 1565
 Leu Tyr Asp Val Val Ser Thr Leu Pro Gln Val Val Met Gly Ser Ser
 1570 1575 1580
 Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn
 1585 1590 1595 1600
 Thr Trp Lys Ser Lys Lys Asn Pro Met Gly Phe Ser Tyr Asp Thr Arg
 1605 1610 1615
 Cys Phe Asp Ser Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu Ser
 1620 1625 1630
 Ile Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Lys
 1635 1640 1645
 Ser Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys
 1650 1655 1660
 Gly Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr
 1665 1670 1675 1680
 Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala
 1685 1690 1695
 Cys Arg Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Asn Ala Ala
 1700 1705 1710
 Gly Leu Val Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala
 1715 1720 1725
 Ser Leu Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro
 1730 1735 1740
 Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys
 1745 1750 1755 1760
 Ser Ser Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr
 1765 1770 1775
 Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu
 1780 1785 1790
 Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met
 1795 1800 1805
 Tyr Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe
 1810 1815 1820
 Ser Ile Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln
 1825 1830 1835 1840
 Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile
 1845 1850 1855
 Ile Glu Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser
 1860 1865 1870
 Pro Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val
 1875 1880 1885
 Pro Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg
 1890 1895 1900
 Leu Leu Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe
 1905 1910 1915 1920
 Asn Trp Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala
 1925 1930 1935

Ser Gln Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly
 1940 1945 1950
 Asp Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu
 1955 1960 1965
 Cys Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn
 1970 1975 1980

Arg
 1985

<210> 2
 <211> 5965
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Non-optimized cDNA sequence encoding SEQ. ID. NO.
 1

<400> 2
 gccacacatgg cgcccatcac ggccctactcc caacagacgc ggggcctact tgggtgcac 60
 atcaatagcc ttacaggccg ggaacaagaac caggtcgagg gagaggttca ggtggtttcc 120
 accgcaacac aatccttctc ggcgacctgc gtcaacggcg tgtgttgagc cgtttaccat 180
 ggtgctggct caaagacactt agccggccca aagggggcaa tcaccagat gacacataat 240
 gtggaccagg acctctcgct ctggcaggcg cccccgggg cgcgttctct gacacatgc 300
 acctgtggca gctcagacct ttacttggct acgagacatg ctgacgtcat tccggtgccc 360
 cggcggggcg acagttaggg gagcctgctc tccccaggc ctgtctctca ctggaaggcg 420
 tcttcgggtg gtccactgct ctgcccttgc gggcacgtcg tgggcattct ccgggtgccc 480
 gtatgcaccc ggggggttgc gaaggcggtg gactttgtgc ccgtagagtc catggaaact 540
 actatgcggg ctccggtctt caccgacaac tcattcccc ccggcgtacc cagctcattt 600
 caagtggccc acctacacgc tcccactggc agcgccaaga gtactaaagt gccggctgca 660
 tatgcagccc aagggtacaa ggtgctcgtc ctcaatcgt ccgttgccgc taccttaggg 720
 ttggggcgct atatgtctaa ggcacacggt attgacccca acatcagaac tggggtaagg 780
 accattacca caggcgcccc cgctcacata tctacctatg gcaagtttct tggcgatggg 840
 ggttgctctg ggggcgctta tgacatcata atatgtgatg agtgccattc aactgactgc 900
 actacaactc tgggcacatg cacagtctcg gaccaagcg agacggctgg agcgcggtct 960
 gtggtgctcg ccaccgctac gctctcgga tcggtcaccg tggcacacc aaacatcgag 1020
 gaggtggccc tgtctaatac tggagagatc ccttctatg gcaagccat cccattgaa 1080
 gccatcaggg ggggaaggca tctcatttct tgtcattcca agaagaagt cgacgagctc 1140
 gcgcgaagcg tgcaggcct cggaaataac gctgtggcgt attaccgggg ctcgatgtg 1200
 tctgcataac caactatcgg agacgtcgtt gtctgtggca gacagcctct gatgacggcg 1260
 ttacgctggg actttgactc agtatcgac tgaacacat gtgtcaccca gcaactcgac 1320
 ttacgcttgg atcccaacct caccattgag acgacagccg tgcctacaaga cgcagtgtgc 1380
 ccgtcgcagc gggcggttag gactgcccgg ggtaggagag gcatctacag gttgtgact 1440
 ccgggagaac ggcctcggg catgttcgat tctcgtgtcc tgtgtgagtg ctatgacgcg 1500
 ggcgtgtcct ggtacagact caccoccccg gagacctcgg ttagggttgcg ggcctacctg 1560
 aacacaacac ggttgccctt ttgccaggac cacttgaggt tctgggagag tctcttcaca 1620
 ggcctcacc acatagatgc acacttctgt tcccagacca agcaggcagg agacaacttc 1680
 cctacctcgg tagcatacca agccacgggt tggccaggg ctcaggcccc accctcatca 1740
 ttggatcaaa tgtggaagtg tctcatcagg ctgaacacct cgcgtcacgg gccaaacccc 1800
 ttgctgtaca gctcgggagc cgtccaaaat gaggtcaccc tcaccacccc cataaccaaa 1860
 tacatcatgg catgcattgc ggtcgacctg gaggtcgtca ctgaccacct ggtgcttggt 1920
 ggcggagctc ttgcactctc gcccgcgtat tgcccgacaa caggcagtggt ggtcatttgt 1980
 ggttagatta tctgttccgg gaggccggct attgttcccg acagggagtt tctctaccag 2040
 gagttcgatg aatggaaga gtgcgctcgc caactccctt acatcgagca gggaaatgag 2100
 ctgcgcgagc aattcaagca gaaagcgctc ggggttactgc aaacagccac caaacaagcg 2160

gaggetgttg	ctcccggtgg	ggagtcocag	tggcgagccc	ttgagacatt	ctggcggaag	2220
atctctatcg	atttctatcg	cgggatacag	tactatagcag	gcttatccac	ctctgctggg	2280
aaaccccgaa	tagcatcatt	gatggcattc	acagcctcta	tcacacagccc	gctcaccacc	2340
caagtagccc	tctctgttaa	catctbtggg	gggtgggtgg	ctgcccaact	cgcccccacc	2400
agcccgctt	cggtcttcgt	ggggcgccgc	atcgccgggtg	cggtctgttg	cagcataggg	2460
ctggggaagg	tgtctctggga	caattctggg	ggttatggag	caggagtggc	cgccgcgcgc	2520
gtgcgctcca	aggtcatagag	cggcgagatg	ccctccacag	aggacctggg	caactctact	2580
ctgcgcattc	tctctctcgtg	cgccctggtc	gtcggggctg	tgtgtgcagc	aatactgggt	2640
cgacacgtgg	gtccggggaga	gggggctgtg	cagtggatga	acgggctgat	agcgttcgcc	2700
tcgcggggta	atcatgttttc	ccccacgcac	tatgtgcctg	agagcgagcg	cgccagcgct	2760
gttactcaga	tctctctccag	ccttaccatc	actcagctgc	tgaaaagggt	ccaccagtgg	2820
attaatgaag	actgcctccac	acgtgttctc	ggctcgtggc	taagggaatg	ttgggactgg	2880
atatgcacgg	tgttgactga	cttcaagacc	tggctccagt	ccaagctcct	gcgcagacta	2940
ccgggagctc	ctttttttct	gtgcacaacg	gggtacaagg	gagttctggc	gggagacggc	3000
atcatgcaaa	ccacctgccc	atgtggagca	cagatcaccc	gacatgtcaa	aaacgggtcc	3060
atgaggatcg	tcgggctcaa	gacctgcagc	aaacgtggc	atggaaacatt	ccccatcaac	3120
gcatacacca	cgggcccccgt	cacacctctc	ccagcgccaa	actattctag	ggcgctgtgg	3180
cggttgccgc	ctgagggagta	cgtggaggtc	acgcgggtgg	gggatttcca	ctactgtgac	3240
ggcatgacca	ctgacaacgt	aaagtgcaca	tgccaggttc	ggcctctcga	attcttccgc	3300
gaggtgggacg	gagtgccggtt	gcacaagttac	gctcccgctg	cgagggcctct	cctacgggag	3360
gaggttacat	tccagggtcgg	gctcaaccaa	tacttggttg	ggtcacagct	accatgcgag	3420
cccgacaacgg	atgtgacagt	gctcaattcc	atgtctacgc	acccctccca	catcacagca	3480
gaaacggccta	agcgtaggtt	ggccaggggg	tctccccctc	ccttggccag	ctcttcagct	3540
agccagctgt	ctgcgccttc	cttgaaggcg	acatgaccta	cccacacatg	ctctccggag	3600
ctgcacctca	tcgaggccaa	cctcctgttg	cggcaggaga	tgggcgggaa	catcacccgc	3660
gttgagtcgg	agaaacaagt	ggtagtctctg	gactcttctg	acccgctctg	agcggagagg	3720
gatgagaggg	aagtatccgt	tcggcgggag	atcctgcgga	aatccaagaa	gttccccgca	3780
gcgatgacca	tctgggcggc	cccggtattac	aacctctcac	tgttagagtc	ctgggaaggac	3840
ccggactacg	tcctctccgt	ggtgcacggg	tgcccgttgc	ccactatcaa	ggccctccca	3900
ataccacctc	cacggagaaa	gaggacggtt	gtcctaacag	agtcctccgt	gtcttctgcc	3960
ttagcggagc	tcgtctactaa	gaccttcggc	agctccgaat	catcgccgtg	cgacagcgcc	4020
acggcgacgc	cccttctctg	cgaggcctcc	gacgacgggt	acaaaggatc	cgacgttgag	4080
tcgtactctc	ccatgcccccc	ccttgagggg	gaaccggggg	accccgatct	cagtgacggg	4140
tcttgtctta	ccgtgagggca	ggaagctagt	gaggatgtcg	tctgctgctc	aatgtctcac	4200
acatggacag	gcgccttgat	cacgccaatg	gctcgggagg	aaagcaagct	gccactcaac	4260
cggttgagca	actcttttgt	gcgccaccat	aaatgtgttt	atgcccaaac	atctccgcag	4320
gcagcctctg	gcgagagaag	ggtcaccttt	gacagactgc	aagtctctga	cgaccactac	4380
cgggacgtgc	tcaaggagat	gaaggcggaag	cgctccacag	ttaaggctaa	actcctatcc	4440
gtagaggaag	ctcgcaagct	gacgcgccca	catctggcca	aatccaaagt	tggctatggg	4500
gtaaaagagc	tcgggaacct	atccagcaag	cgcgtaaac	acatccaact	cgtgtgggaag	4560
gactctgttg	agacacactg	gacaccaatt	gacaccacca	tcatggccaa	aaatgaggtt	4620
tctgtgtctc	aaccagagaa	aggaggccgt	aagccagccc	gcttatctgt	atttcccaat	4680
ctgggagtgc	gttatgcca	gaagatggcc	ctctatgatg	tgggtctcac	cctctctcag	4740
gtcgtgatgg	gtctctcata	cggattccag	tactctctct	ggcagcgagt	cgagttctcg	4800
gtgtaactct	ggaaatcaaa	gaaaaacccc	atggggcttt	catatgacac	tcgtgttttc	4860
gactccaacg	tcacccagaaa	gcacatccgt	gttgaggagt	caatttacc	atgttgtgac	4920
ttggccccgc	aagccagagca	ggccataaaa	tcgtccacag	acggcgctta	tatcgggggt	4980
cctcttgacta	attccaaggg	cgagaactgc	ggttatcgcc	gctgcggcgc	gagcgccgtg	5040
ctgacgacta	gctcggttaa	cacctctaca	tgttaactga	agggcctctg	agcctgtcga	5100
cttcgggaagc	tcagggactg	cacgagtctc	gtgaacggcg	cgccgcttgt	cgttatctgt	5160
gaaagcgccg	gaacccaaga	ggagcgccgc	agcctacagag	tcttccacgga	ggctatgact	5220
aggtactctg	cccccccggc	ggacccggcc	caaccagaaat	acgacttga	gctgataaca	5280
tcagtgtctc	ccaaatgtgtc	ggtcggccac	gatgcacacg	gcgaagggtg	gtactacctc	5340
acccgctgac	ccaccacccc	ctcgcacgg	gctgctgggg	aaacagctag	acacactcca	5400
gttaactctc	ggctaggcaaa	catctatcat	tatgcgccca	ctttgtgggc	aaggatgat	5460

ctgatgactc	acttettctc	cattcttctc	gcacaggagc	aacttgaaaa	agccctggac	5520
tgccagatct	acgggggctg	ttactccatt	gagccacttg	acctactcca	gatcattgaa	5580
gcactccatg	gccttagcgc	attttcactc	catagttact	ctccaggtga	gatcaatagg	5640
gtggcttcat	gcctcaggaa	acttggtgta	ccaccctctg	gagttctggag	acatcggggc	5700
aggagcgtcc	gcgcctaggct	actgtcccag	ggggggaggg	cgcccaactg	tggcaagtac	5760
ctcttcaact	gggcagtgaa	gaccaaaact	aaactcactc	caatccccgc	tgctgtccag	5820
ctggacttgc	ccggactgttt	cgtgtcgtgt	tacagcgggg	gagacatata	tcacagctcg	5880
ttctgtgcc	gaccgccgtg	gttcatgttg	tgccatactc	tactttctgt	aggggttaggc	5940
attcactcgc	tccccaacgc	ataaaa				5965

<210> 3

<211> 5965

<212> DNA

<213> Artificial Sequence

<220>

<223> Optimized cDNA encoding SEQ ID NO: 1

<400> 3

gcacacatgg	cccccatcac	cgccctacag	cagcagaccc	gcggcctgct	gggctgcac	60
atccacagcc	tgaccggcgc	cgacaaagac	caggtggagg	gcgaggtgca	gggtgtgagc	120
acgcgcaccc	agagcttctc	ggccacactgc	gtgaacggcg	tggtgtggag	cgtgtaccac	180
ggcgcccgca	caagtagccct	ggccggcccc	aagggcccca	tcaccagat	gtacacccaac	240
gtggaccagg	acctggtggg	ctggcaggcc	ccccccggcg	ccgcagacct	gacccctcgc	300
acctgcggca	gcagcgacct	gtacctgggt	accgcgccag	cgacgtgat	ccccgtgcgc	360
cgccgcggcg	acagcgcggg	cagcctgctg	agcccccgcc	ccgtgagcta	cctgaagggc	420
agcagcggcg	gccccctgct	gtgccccagc	ggccacggcg	tggtgatctt	ccgcgcggcc	480
gtgtgcaccc	gcggctgtgc	caaggccgtg	gacttcgtgc	ccgtggagag	catggagacc	540
acctgcgcga	gccccctgct	caccgacaac	agcagccccc	ccgcgtgccc	ccagagcttc	600
caggtggccc	acctgcacgc	ccccacggcg	agcggcaaga	gcaccaaggt	gcccgcggcc	660
tacgcggccc	agggtacaaa	ggtgctgggt	ctgaacccca	gcgtggccgc	cacctggggc	720
ttcgccgccc	acatgagcaa	ggccccacgc	atcgacccca	acatccgcac	cgccgtggcg	780
accatcacca	ccgggcggcc	cgtgacctac	agcactatcg	gcaagttcct	ggccgacggc	840
ggctgcagcg	gcgggcgccta	cgacatctac	atctggcgag	agtgccacag	caccgcagac	900
acacacatcc	tggggatcgg	caaccgtgctg	gaccagggcg	agaccgcggg	gcgccgcctg	960
gtggtgtcgt	ccaccgccac	cccccccgcg	agcgtgacgc	tgccccaccc	caacatcgag	1020
gaggtgcgcc	tggccaacac	cggcgagatc	ccctctctacg	gcaaggccat	ccccctcag	1080
gcacatcccg	gcgggcgccta	cctgatcttc	tgccacagca	agaagaaagt	cgacgagctg	1140
gcgcgaagc	tgagcggcct	gggcatcaac	gcctggcctc	actaccgcgg	cctggacgtg	1200
agcgtgatcc	ccacctcgag	cgacgtgggt	gtgggtggca	ccgacgcctc	gatgacggcg	1260
tacacggcg	acctcgacag	cgtgatcgac	tgcacaacct	gcgtgacca	gaccgtggaac	1320
ttacagctgg	accccacctt	cacctcgag	accacacacc	tgccccagga	gcgcctgagc	1380
cgacgcagcg	gcgcggcgcg	caccggccgc	ggccgcgcgc	gcactatccg	cttctgacc	1440
cccgcgagcg	gcgccagcgg	catgtctcag	agcagcgtgc	tggtgcagtg	ctacgacgcc	1500
ggctgcgccc	ggtacagact	gacccccgcg	gagacacagc	tgccctcgcg	cgccctacct	1560
aacacccccg	gcctgcgccc	gtgccaggac	cacctggagt	tctgggagag	cgtgttccac	1620
ggcctgaccc	acattcgacg	ccacttctct	agccagacca	agcaggccgg	cgacacactc	1680
ccctacctgg	tggtctacca	ggccaccgtg	tgcccccggc	ccccggcccc	cccccccgag	1740
tgccaccaga	tggtgaagtg	ctgatccgc	ctgaagccca	ccctgcacgg	ccccaccccc	1800
ctgtgtatcc	ggctgggcgc	cgtgcagaac	gaggtgacct	tgacccaccc	catcaccaag	1860
tacatcatgg	cctgcgatgg	gcgcgaactc	gaggtgtgga	ccagcacctg	gggtctgggt	1920
ggcggcgtgc	ggcgccgccc	ggccgcctac	tgctgacca	ccggcagcgt	gggtatcgtg	1980
ggcgcgatca	tctgtacggg	ccgcgccccc	atcgtgcccg	accgcagagt	cctgtaccag	2040
gagttcgagc	agatggagga	gtgcgcagcg	cacctgcctc	acatcgagca	gggcattgaag	2100
ctggccgagc	agttaagaa	gaaggccctg	ggcctgctgc	agaccgccac	caagcagggc	2160

gaggcgccgc	cccccggtg	ggagagcaag	tggcgccgc	tggagacctt	ctggggccaag	2220
cacatgtggg	acttcatcag	cggcatccag	tacctggccg	gcttgagcac	ctccgcggcg	2280
aaaccccgcca	tcggccagct	gatggccctc	accgccagca	tcaccagccc	ctctgaccacc	2340
cagagcacc	gtggttcaa	catcctgggc	ggctgggtgg	cgcccgagct	ggccccccccc	2400
agcgccgcca	gcgccttcgt	ggggcgccgc	atcgccggcg	cgccgctggg	cagcatctggc	2460
ctggggcaag	ctctgggtgga	catcctggcc	ggctacggcg	cgccgctggc	cgccgcccgt	2520
gtggcccttca	aggtgatag	cggcgagatg	accagaccgc	aggacctggt	gaacctgctg	2580
cccgccatcc	ttagccccgc	cgccctgggt	gtggggctgg	ttgtcgccgc	catctgtcgc	2640
cgccacgtgg	gcgccggcga	ggggcgccgt	cagtggatga	accgctgat	cgccctgcgc	2700
agcccgccga	accacgttag	ccccaccacc	tactgtccgc	agagcgagcg	cgcccgccgc	2760
gtgaccacga	ctctgagcag	ccctgaccatc	accagctgc	tgaagcgcc	gcaccagtgg	2820
atcaacagag	actgcagcac	ccctgcagcg	ggcagctggc	tgcgcgact	gtgggactgg	2880
atctgcacgc	tgttgaccga	ctcaagacc	tggtgcaga	gcaagctgt	cgccagctgg	2940
cccgccgtgc	ctcttctcag	ctgcacagcg	ggctacaagg	ggtgtggcg	cgccgacggc	3000
atcatgcaga	ccacctgccc	ctgcggcgcc	cagatacccg	gccacgtgaa	gaacggcagc	3060
atgcgcacg	tgggccccaa	gacctgcagc	aacacctggc	acggcacctt	ccccatcaac	3120
gcctacacca	cgcccccctg	caccccccag	cccgccccca	actacagccc	cgccctgtgg	3180
cgctgtggcg	ccgagagtag	ctgtggaggtg	acccgctgtg	cgacttccca	ctacgtgacc	3240
ggcatgacca	cgcacaagct	gaagtgcgcc	tgccaggtgc	cgccccccga	gttcttacc	3300
gaggtgtgagc	gctgtgcctg	gcaccgcctac	gcccccgctc	gcccgcctcc	gctcgctgag	3360
gaggtgtacct	tcaggtggg	ctgaaccag	tacctgggtg	gcagccagct	gctcgctgag	3420
cccgagccgc	agctggccgt	gctgaccagc	atgctgaccg	acccccagca	catcacccgc	3480
gagacggcca	agcgcccgct	ggccccggcg	agccccccca	gcttgccca	cagcagcgcc	3540
agccagctga	gcgccccccg	ctgaaggcc	actgcagcca	ccacacagct	gagccccagc	3600
gcgcagctga	tgcaggccaa	ctgtctgtgg	cgccaggaga	tggcgggcaa	catcacccgc	3660
gtggagagcg	agaaacaagt	ggtggtgctg	gacagcttgc	acccccctgc	cgccgagagc	3720
gacgagcgcg	aggtgagcgt	gccccggcag	atcctgcgca	agagcaagaa	gttccccgc	3780
cccatgccca	cttggggccg	ccccgactac	aacccccccc	tgtggagag	ctgggaaggac	3840
ccccgactac	tgcccccggt	ggtgcacggc	tgccccctgc	cccccatcaa	ggcccccccc	3900
atcccccccc	cccccgccaa	cgccacgtgc	gtgctgaccg	agagcagcgt	gagcagcgcc	3960
ctcgccgagc	tgggccaccaa	gaccttcggc	agcagcgaga	gcagcgccgt	gagcagcgcc	4020
accgccaccg	ccctgcccga	ccaggccagc	gacgacggcg	acacgggacg	cgacgtggag	4080
agctacagca	gctgcccccc	ctggagggcg	gagccccggc	acccccgact	gagcagcgcc	4140
agctggagca	ccgtgagcga	ggaggccagc	gaggacgtgg	tgtgtgagc	catgagctac	4200
accctggaccg	gcgcctctgat	caccccctgc	gcgcgcgagg	agagcaagct	gccccatcaac	4260
gcctcgagca	acagcctctg	cgccccacc	aatatgggtg	acgccaccac	cagcccgagc	4320
gcgcgctgc	gcagagaagaa	ggtgaccttc	gaccgcctgc	aggtgctgga	gcaccactac	4380
cgccagctgc	tgaaggagat	gaaggccaa	gccagaccgc	tgaaggccaa	gctgctgagc	4440
gtgagagagc	ctccaagct	gacccccccc	cacagcgcca	agagcaagtt	cgcttacggc	4500
gccaagagc	tgcgcaact	gagcagcaag	gcctgaaacc	actccacag	cgtgtggag	4560
gacctctgag	agggacacgt	gacccccatc	gacacaccca	tcattggccaa	gacagaggtg	4620
ttctgtgtgc	agccccagaa	ggggcgccgc	aagccccccc	gctgtactgt	gttccccgac	4680
ctggcgctgc	cggtgtgcga	gaagatggcc	ctgtacgagc	tggtgagcac	ctccccccag	4740
gtgggtgctg	gcagcagcta	cggttccag	tacagccccc	gccagcgctg	gaggttccgt	4800
gtgaacacct	ggaagagcaa	gaagaaacccc	atgggtcttca	gctacacac	cgctgtcttc	4860
gcagcagccg	tgcacagaa	ggacatccgc	gtggagagga	gcatctacca	gtgctgcgac	4920
ctcgcccccg	agggcccgca	ggccatcaag	agcttgacgc	agcgctgtga	catcgccggc	4980
ccctcgacca	acagacaagg	ccagaactgc	ggctacacgc	agggcagcgc	cagcgcgctg	5040
ctgacaccaca	gctgtggcga	caacctgacc	tgctacacga	agggcagcgc	cgcttgcgc	5100
gcgcgcaagc	tgacaggact	cacctgctg	gtgaagcgcc	cgccctgggt	gggtgactgc	5160
gagtagcgcc	gcacccagga	gcagcgccgc	agcttcggcg	tggttcaacga	ggccatgacc	5220
ctgtacagcg	ccccccccgg	gcaccccccc	cagccccagc	acgacctgga	gctgataacc	5280
agctgcagca	gcacgttag	ctgggcccac	gacgccaagc	gcaagcgctg	gtacttacctg	5340
accgcgaccc	ccacaccccc	ctggcgccgc	gcgcctggg	agacggcccc	ccacaccccc	5400
gtgaacagct	gcttgggcaa	catcatcatg	tacgccccca	ccctgtgggg	ccgcatgac	5460

ctgatgaccc	actttcttcag	catcctgctg	gcccaggagc	agctggagaa	ggccctggac	5520
tgccagatct	acggggcgctg	ctacagcctc	gagcccccctg	acctggccca	gatcatcgag	5580
cgcctgcacg	gcccagagcg	cttcagccctg	cacagctaca	gccccggcga	gatcaaccgc	5640
gtggccagct	gcctgcgcga	gctggggcgtg	cccccccctg	gcgtgtggcg	ccaccgcgcc	5700
cgacgctgct	gcgcccgccct	gctgagccag	ggcgcccgcg	gcgccacctg	cggcaagctac	5760
ctgtttcaact	ggggcctgtaa	gaccaagctg	aagctgaccc	ccatcccccg	cgccagccag	5820
ctggaacctga	gcggctgggtt	cgtggccggc	tacagcgccg	gcacacatcta	ccacagcctg	5880
agccgcgccc	gcccccgctg	gttccatgctg	tgccctgctgc	tgctgagcgt	ggggcgtgggc	5940
atctaccctgc	tgcccaaccg	ctaaa				5965

<210> 4

<211> 37090

<212> DNA

<213> Artificial Sequence

<220>

<223> MRKAd6-NSmut nucleic acid

<400> 4

catcatcaat	aatataacct	atcttggtatt	gaagccaata	tgataatgag	gggggtggagt	60
ttgtgacgtg	gcgcggggcg	tgggaaacggg	gcgggtgacg	tagtagtggt	gcggaagtgt	120
gactgttgcaa	gtgtgcgcga	acacatgtaa	gcgacggatg	tgccaaaagt	gacgtttttg	180
gtgtgctggcg	gtgtacacag	gaagtgtacaa	ttttcgcgcg	gttttagcgc	gatgtttgtg	240
taaaattcggt	cgtaaccgag	taagatttgg	ccattttcgc	gggaaaactg	ataagaggga	300
agtgaattct	gaataattttt	gtgttactca	tacgcgttaa	tatttgtcta	gggccgcggg	360
gactttgaccc	gtttacgttg	agactcgccc	aggtgttttt	ctcagggtgt	ttccgcgttc	420
cggtgtcaaat	ttggcgtttt	attattatag	gcggccgcga	tccattgcat	acgtttgtatc	480
catatcataa	tatgtacatt	tattttggct	catgtccaac	attaccgccca	tgttgacatt	540
gattattgac	tagttatttaa	tagtaatacaa	ttacggggtc	attagttcat	agcccatata	600
tggaagttccg	cgttacataa	ctacaggtaa	atggcccgcc	tggtgacgc	cccaacgacc	660
ccgcgccatt	gacgtcaata	atgacgtatg	ttcccatagt	aacgcccaata	gggactttcc	720
attgacgtca	atgggtggag	tatttacgtg	aaactgccca	cttggcagta	catcaagtgt	780
atcatatgct	aagtacgccc	cctattgacg	tcaattgacg	taaatggccc	gcttggcatt	840
attgccagta	catgacctta	tgggactttc	ctacttgcca	gtacatctac	gtatttagtca	900
tcgtctattac	catgtgtgatg	cgtttttggc	agtaacatca	tgggcgtgga	tagcgttttg	960
actccagggg	atttccaaat	ctccacccca	ttgacgtcaa	tgggagtgtt	ttttggcacc	1020
aaaatacaacg	ggactttcca	aaatgtogta	acaactccgc	cccatgacgc	caaatggggc	1080
gtaggcgtgt	acggtggggg	gtctatataa	cgagagctcg	tttagtgaac	cgtcagatcg	1140
ctcggaagcg	catcccaacg	tgttttgacc	tcctatagaag	acacccggac	cgatccagcc	1200
ttccgcggcg	ggagacggctg	atgggaacgc	ggattccccg	tgccaaagat	gagatctgcc	1260
acacatgctgc	catcacgcgg	ctactcccaa	cagacgcggg	gctacttggg	ttgcatctac	1320
actagcctta	cagccgggga	caagaaccag	gtcggaggag	aggttcaggt	ggtttccacc	1380
gcaacacata	ctttccctggc	gaactcgctc	aacggcgctg	gttggagcgt	ttacatcgtg	1440
ctggtgtcaa	agacccttagc	cgcccgcaag	gggccaatca	cccagatgta	caactattgt	1500
gaccaggacc	tcgtcggctg	gcaggcgccc	ccggggcgcg	gttccctgac	accatgcacc	1560
tggtggacgt	cagacccttta	cttgggtcaag	agacatgctg	agctcatctc	gggtgcgcgg	1620
cggggcgaca	gtaggggggg	cctgcctctc	cccaggccctg	tctctactct	gaagggtcct	1680
tcgggtgtgtc	cactgctctg	cccttcgggg	cacgtgtgtg	gcattcttcg	ggctgcgcta	1740
tgaccgccgg	gggttgcgaa	ggcggtggac	tttgtccccc	tagagtccat	ggaaactact	1800
atgcggtctc	cgtgtcttcc	ggacaactca	tcctccccgc	cgtacacgca	gtcatttcaa	1860
gtggcccaacc	tacacgctcc	cactgcgcgc	ggcaagagta	ctaaaagtgc	ggctgcatac	1920
gcagcccaag	gtgtacaaggt	gctcgtctc	aatccgttcc	tgctgcctac	cttagagctt	1980
ggggcgtata	gttctaagcg	acacggtatt	gaccccaaca	tcagaactgc	ggtaaggacc	2040
attaccagag	gcgcgccctg	cacatactct	acctatggca	agttctctgc	cgtatggtagt	2100
tgctctgggg	gcgcttatga	catcataata	tgtgatgagt	gcatctcaac	tgactcgact	2160

acaatcttgg	gcatacggac	agtcctggac	caagcggaga	cggctggagc	gcggtctgtc	2220
gtgctcgcca	ccgctacgac	tccgggatac	gtccacgtgc	cacacccaaa	catcgaggag	2280
gtggccctgt	ctaatactgg	agagatcccc	ttctatggca	aagccatccc	catggaagcc	2340
atacgggggg	gaaggcatct	catttttgtg	cattccaaga	agaagtgcca	cgagctcgcc	2400
cgaaagctgt	caggctccgg	aatcaacgtc	gtggcggtatt	acggggggct	cgatgtgtcc	2460
gtcataccaa	ctatcggaga	cgctgtgtgc	gtggcaacag	acgctctgat	gacggcttat	2520
acgggggact	ttgactcagt	gatcgactgt	aacacatgtg	tcaaccagac	agtcgacttc	2580
agcttggatc	ccacctccac	catttgagac	acgacccgtc	ctcaagacgc	agtgtcgccg	2640
tcgcagcggc	ggggtaggac	tggcagggtg	aggagagcca	tctacaggtt	tgtgactccg	2700
ggagaaacgg	ctctggggcat	gttcgattcc	tgggtcctgt	gtgagtctca	tgaacggggc	2760
tgtctctggg	acgagctcac	ccccgccgag	acctcggtta	ggttgccggc	ctacctgaac	2820
acacccaggt	tgcccgcttg	ccaggacccac	ctggagtctc	gggagagtgt	cttcacaggc	2880
ctcaccacca	tagatgcaca	cttcttgtcc	cagaccaagc	aggcaggaga	caacttcccc	2940
taccttgtag	cataccaagc	cacggtgtgc	gccagggtct	agggccccac	tcctcatbgt	3000
gatacaaatgt	ggaagtgtct	catacggctg	aaacctaacg	tgacggggcc	aaacaccttg	3060
ctgtacaggc	tgggagccgt	ccaaaatggag	gtcaccctca	cccccccatc	aaacaaatat	3120
atcatggcat	gcactgtcgc	tgaacctggag	gtcgtcaact	gcacctgggt	gctgggtggc	3180
ggagtcctgt	cagctctcgg	cgcgtattgc	ctgacaacag	cgagtgtggt	catgtgtggg	3240
aggattatct	tgtccgggag	gcggcgtatt	gttcccgaca	gggagtttct	ctcacaggag	3300
ctgtatgaat	tgggaagatg	cgctctcgac	ctcccttaca	tcgacagagg	aatgcagctc	3360
ccgcagcaat	tcaagcagaa	agcgctcggg	ttactgcmaa	cagccacca	acacggcgag	3420
gctctgtctc	ccgtgtgtga	gtccaaatgg	cagccctbgt	agacattctg	ggcgagagac	3480
atgtgtgaatt	tcatcagcgg	gatacagctc	ttagcagcgt	tatccactct	gctctgggaac	3540
ccgcacaatg	catcatgtat	ggcatctaca	gcctctatca	ccagcccgct	caccacccaa	3600
agttacactc	tgtttaacat	cttggggggg	tgggtggctg	cccaactcgc	cccccccgac	3660
gcgcctcggg	cttctgtgtg	cgccggcctc	ggcgtgtcgg	ctgtgtggag	cataggcctt	3720
gggaaggtgc	tgttgacat	tctggcgggt	tatggagcag	gagtgccggg	cgctctctgt	3780
gcc tccaagg	tcattgagcg	cgagatgcc	tccaccgagg	acctgtgtcaa	tctacttctc	3840
gccatctctc	ctcctggcgc	cttggtcgtc	ggggtcgtgt	gtgcagcaat	actgcgtcga	3900
cacgtgggtc	cgggagaggg	ggctgtgcag	tggatgaacc	ggctgatagc	gttcgcctcg	3960
cgggtaatac	atgtttcccc	cacgcactat	gtgcctgaga	gcgacgcgcg	agcgcggtgt	4020
actcagatcc	tctccagcct	taccatcatc	cagctgccta	aaaggctcca	ccagtggatt	4080
aatgaagact	gtctccacac	gtgttccggc	tctgtggctaa	gggatgtttg	ggactggata	4140
tgcacgggtt	tgaactgact	caagacctgc	ctccagtcca	agctctcgcc	cgagctacgc	4200
ggagtcctct	ttttctcgtg	ccaacggggg	tacaaggggg	tctggcgggg	gacggcgatc	4260
atgcaaacca	ctcgcctcat	tggagcacag	atcaccggac	atgtcaaaaa	cggttccatg	4320
aggatcgtcg	ggcctaagac	ctgcagcaac	acgtggcctg	gaacattccc	catcaacgca	4380
tacaccacgg	gcocctgcac	acctctccca	gcgccaaact	attctagggc	gctgtggcgg	4440
gtggcccgct	aggagtacgt	ggaggtcacg	cggttggggg	atttccacta	cgtagcgggc	4500
atgaccactg	acaacgtaaa	gtgcccactg	caggttccgg	ctcctgaatt	ctcacgggag	4560
gtgacgggag	tgcggttgca	caggtacgct	ccggcgctga	ggcctctcct	acgggaggag	4620
gttcaatccc	agctcggggt	caaccaatc	ctggttgggt	cacagctacc	atgcgagccc	4680
gaaccgggat	tagcagtgct	cacttccatg	ctcacgcacc	ctccccacat	cacagcagaa	4740
acggctaaag	tgaagttggc	cagggggtct	ccccctctct	tggccagctc	tccagctagc	4800
cagttgtctg	cgctctctct	gaaggcgaca	tgcatacccc	acatgtctc	tcgggacgct	4860
gacctcactg	aggccaacct	ctgttggcgg	caggagatgt	gcgggaaact	caccggcggt	4920
ggctcggaga	acaaggttgt	agtcctggac	tctttcgacc	cgcttcggag	ggaggaggat	4980
gagagggaag	tatccgttcc	ggcggagatc	ctgcggaaat	ccaagaagt	ccccgcagcg	5040
atgcccattc	ggcgcgcccc	ggattacaac	ctctcactgt	tagagtctct	gaaggaccgc	5100
gcactacgtc	ctccgggtgt	gcacgggtgc	ccgttgcac	ctatcaagge	ccctccaata	5160
gccactccac	ggagaaagag	gacggtgtgc	ctaacagagt	cctcgtgtc	tctgacctta	5220
cgagagctcg	ctacaaagac	cttcggcagc	tccgaatcat	cgccgtcgca	cagcgccacg	5280
gcacccgcc	tctctgacca	ggcctccgac	gacggtgaca	aggagtcgca	cgttgagctg	5340
tactctccca	tgcccccctc	tgaaggggaa	ccgggggacc	ccgatctcag	tgcagggtct	5400
tggctctacg	tgaagcagga	agctagtgcg	gatgtcgtct	gctgctcaat	gtctcacaca	5460

tggacaggcg	ccttgatcac	gccatgcgct	gcggaggaaa	gcaagctgcc	catcaacgcg	5520
ttggagcaact	ctttgtctgcg	ccaccataaac	atgggtttatg	ccacaacaatc	tcgcagcgca	5580
ggcctgcggcg	agaagaaggt	ccaccttgac	agactgcaag	tctggacga	ccactacccg	5640
gcagctgctca	aggagatgaa	ggcgaaggcg	tcacacgtta	aggctaaact	ccatccogta	5700
gaggaagcctc	gcaagctgac	gcccccacat	tcggccaaat	ccaagtttgg	ctatggggca	5760
aaggacgtcc	ggaactctatc	cagcaaggcc	gttaaccaca	tccactccgt	gtggaaggac	5820
ttgtctggaag	acactgtgac	accaattgac	accacatac	tgccaaaaaa	tcagggttttc	5880
ttgtgtccaac	caggaagaag	aggcgttaag	ccagcccgc	ttatcgtatt	cccagatctg	5940
ggagtcctgtg	tatgcgagaa	gatggccctc	tatgatgtgt	tctccaccc	tctcaaggtc	6000
gtgatgggct	ccctacacg	attccagtac	tctcctggcg	agcgagtoga	gttccgtggt	6060
aatacctggga	aatcaaaagaa	aaaccccatg	ggcttttcat	atgacactcg	ctgtttcgac	6120
tcacagctca	ccgagaacga	catccgtgtt	gaggagtcaa	tttaccaatg	ttgtgacttg	6180
gcccccgaa	cagacacggc	cataaaaatc	ctcacagagc	ggcttttat	tcagggttct	6240
ctgactaat	caaaaaggga	gaactcggtg	tatcgccgtg	gcgcgcgcg	cgccgtgtcg	6300
acgactagct	gcggttaacac	ctcacatgt	tacttgaagg	ccctctgcagc	ctgtcgagct	6360
gcggaagctcc	aggactgcac	gatgtctgtg	aacgcccgcg	gccttgtctg	tatctgtgaa	6420
agcgccgggaa	cccaagagga	cgcggcgagc	ctacgagtct	tcacggagcg	tatgactagc	6480
tactctgcc	ccccggggga	ccgcgcccaa	ccagaatatc	acttggagct	gataacatca	6540
tgttctcca	atgtgtctgt	gcgccacgat	gcactcaggca	aaagggttga	ctactctacc	6600
cggtatccca	ccaccocctc	gcacggggct	gcgtgggaaa	cagctagaca	caactccagt	6660
aaactcctggc	taggcaacat	tatcatgtat	gcgcccaact	ttgtgggcaag	gatgatctct	6720
atgactcaact	tcttctccat	ctctctagca	caggagcaac	ttgaaaaagc	ccctggactcg	6780
cagatctcacg	ggcctgttta	ctccatttag	ccacttgacc	tactcagat	catgaaagca	6840
ctccattggcc	ttagcgcatt	ttcactccat	agttactctc	caggtgagat	caatagggtg	6900
gctctatgcc	tcaggaaaact	tggggtaacca	cccttgcgag	cttggaagaca	tcgggccagg	6960
agcgtcccg	ctaggctact	gtcccagggg	gggaggccgc	ccacttgtgt	caagtacctc	7020
ttcaactggg	cagtgaagac	caaaactcaa	ctcactccaa	tcocggctgc	gtcccagctg	7080
gacttctgcg	ccgtgtctgt	tgctggttac	agcgggggag	acatatatca	cagcctgtct	7140
cgtcccgac	cccgtgtgtt	catgctgtgc	ctactcttac	tttctgtagg	ggtaggcatc	7200
tactctgcc	cccaacggta	aatctagagc	tgtgcttct	agttgccagc	catctgttgt	7260
ttgcctctcc	cccggtcctt	cccttgacct	ggaaggtgcc	actcccactg	tcctttccca	7320
ataaaatgag	gaaattgcac	cgcaattgtct	gagtaggtgt	catctctatt	tcgggggggtg	7380
ggtggggcag	gcacgacaag	gggaggaattg	ggaaacaaat	agcaggcatg	ctggggagtc	7440
ggtgggctct	atggccgatc	ggcgcccgct	actgaaatgt	gtggggcggtg	cttaagggtg	7500
ggaaaagaata	tataaagttg	gggtcttatg	tagtttttga	tctgttttgc	agcagccgcc	7560
gcgcacatga	gcaccaactc	gtttgtgaga	agcatttgtga	gcctcatatt	gacaacgcgc	7620
atgcccocat	ggggccgggt	gcgtcagaat	gtgatgggct	ccagcatgga	gtgtgcgcc	7680
gtcctgcgcc	caaaactcac	tacttgacc	tacgagaacg	tgtctggaac	gcctgtgga	7740
actgcagctc	ccgcgcgcgc	ttcagccgct	gcagccactc	cccgccggat	tgtgactgac	7800
tttgcctctt	tgagccgct	tgcaagcagt	gttcatctcc	gttcactgac	cccgatgac	7860
aagttagcgg	ctcttttggc	acaattggat	tctttgaccc	gggaacttaa	tgtgtttct	7920
cagcagctgt	tgagatctgc	ccagcaggtt	tctgccttca	aggttctctc	ccctcccaat	7980
cggttttaaa	acataaaata	aaaaccagac	tctgttttga	tttggatcaa	gcaagtgtct	8040
tgtctctctt	atttaggggt	tttgccgcg	cggtaggccc	gggaccagcg	gtctcggctg	8100
ttgaaggctc	tgtgtatctt	ttccaggagc	tgttaaaagt	gactctggat	gttcagatc	8160
atgggcaata	gccgctctct	gggttgagg	tagcacactc	ccagagcttc	atgtcgggg	8220
gtggtgttgt	agatgatcca	gtcgtagcag	gagcgtggg	cgtggtgcct	aaaaatgtct	8280
ttcaagtga	gtgtgatctg	caggggcagg	cccttggtgt	aagtgtttac	aaagcgggtta	8340
agctgggagt	gggtcatatc	tgggggatgt	agatgcactc	tggactgtat	tttttaggtg	8400
gctatgtctc	cagccatact	ctccggggga	ttcatgtttc	gcagaaccac	cagcacagtg	8460
tatccggtgc	acttgggaaa	tttgtcatgt	agcttagaag	gaaatgcgtg	gaagaaactgt	8520
gagacgcctc	gttgacctcc	aagattttcc	atgcattcgt	ccataatgat	ggcatacgtg	8580
ccacggcgcg	cggtcgtggc	gaagataatt	ctgggataac	taacgtcata	gttgtgttcc	8640
aggatagat	gctcatagcc	catttttaca	aagcgcggcg	ggagggtgcc	agactcggtg	8700
ataatggttc	catccgcgcc	aggggcgtag	ttaccctcac	agatttgtat	ttcccaagct	8760

ttgagttccag	atgggggggat	catgtctacc	tgcggggcgca	tgaagaaaac	ggtttccggg	8820
gtaggggaga	tcaagtggga	agaagcagg	ttcctgagca	gctgcgactt	acgcgagccg	8880
gtggggccgt	aaatcacacc	tattaccgcc	tgcaactcgt	agttaagaga	gctcgcagtc	8940
cgtctacc	ttagcagggg	ggccacttcg	ttaagcatgt	ccctgactcg	catgttttcc	9000
cgtaccaaat	ccgcgcgaag	ggcgtcgccg	cccgagcata	gcagttcttg	caaggaagca	9060
aagtttttca	acggtttgag	acgcgtccgc	gtagccatgc	ttttgagcgt	ttgaccaagc	9120
agttccagg	gtgtccacag	ctcgtgcacc	tgtctctacg	catctcgatc	cagcatatct	9180
ccctgtttcg	cggtttgggg	cggctttcgc	tgtacggcag	tagtcgtgtc	tcgtccagac	9240
ggggccaggt	catgttcttc	cacgggcgca	gggttcctcg	cagcgtatgc	tgggtcacgg	9300
tgaaggggtg	cgctccgggc	tgcgcgtcgt	ccaggggtcg	cttgaggctg	gtcctgtctg	9360
tgtctgaagc	ctgcgcgtct	tgcgcctcgg	cgtcggccag	gtagcatatt	accatggtgt	9420
catagtccag	ccctcccgcg	gcgtggccct	tggcgcgcag	cttgcccttg	gaggaggcgc	9480
cgcacagagg	gcagtgcaga	cttttgagg	cgtagagctt	gggcgcgaga	aataccgatt	9540
cgggggagta	ggcatcccg	ccgcagccgc	cgcagacggt	ctcgcatlcc	acgagccagg	9600
tgaactctgg	ccgttcgggg	tcaaaaacca	ggtttccccc	atgctttttg	atgcgtttct	9660
tacctctagt	ttccatgagc	cgtgtccacc	gctcgggtac	gaaaaggctg	tcgttgtccc	9720
cgtatacaga	cttgagagcg	ctgtcctcga	gcgggtgtcc	gcggtcctcc	tgcattagaa	9780
actccgacca	ctctgagagc	aaggctcgcg	tccaggccca	cacgaaggag	gctaattggg	9840
aggggtagcg	gtcgttgttc	actagggggg	ccactcgtcc	cagggtgtga	agcacacatg	9900
cgccctcttc	ggcatcaagg	aagggtgatg	gtttataggt	gtaggccacg	tgaccgggtg	9960
tgtctgaagg	ggggctataa	aaggggttgg	gggcgcgttc	gtcctcactc	ttcttcgcga	10020
cgcgtctcgc	gagggccagc	tgttggggtg	agtactccct	ctcaaaagcg	ggcatgactt	10080
cgctcgtaa	attgtcagtt	tccaaaagcg	agggagattt	gatattcaac	tgcccgcggg	10140
tgatgccttt	gaggggtggc	gcgtccatct	ggtcagaaaa	gacactcttt	ttgttgtcaa	10200
gcttgggggc	aaacgaccgc	tagagggcgt	tggacagcaa	cttgccgatg	gagccgaggg	10260
tttgtttttt	gtcgcgatcg	gcgcgctcct	tggccgcgat	gttttagctc	acgtattcgc	10320
gcgcacagca	ccgcctatcg	ggaagacgg	tgggtgcgtc	ctcgggcact	aggtgcacgc	10380
gcacacccgc	gttgtgcagg	gtgacaaggt	caacgtcgtt	ggctacctct	ccgcgtaggc	10440
gctcgttggt	ccagcagagg	cggccgcctc	tgcgcgagca	gaatggcgtt	agtgggtcta	10500
gctgcgtctc	tgcggggggg	tctgcgtcca	cggtaaaagc	cccgggcagc	agggcgcgtt	10560
cgaagtagtc	tatcttgcct	ccttgcaagt	ctagcgcctg	ctgccatgcg	cggcgcggca	10620
gcgcgcgtc	gtatgggttg	agtggggggc	cccatggcat	ggggtgggtg	agccgggagg	10680
gtagtagtcc	gcaaatgtcg	taaaagtaga	ggggctctct	gagtattcca	agatatgtag	10740
gttagcatct	tccaccgcgc	atgctggcgc	gcacgtaatc	gtatagtctc	tgcgagggag	10800
cggagggttc	gggaccgagg	ttgtcacggg	cggcgtcctc	tgtcgggaag	actatctgcc	10860
tgaagtggc	atgtgagttg	gatgatatgt	ttggacgctg	gaagcaggtt	aagctggcgt	10920
ctctgagacc	taccgcgtca	cgcacgaagg	agggctagga	gtcgcgcagc	tgtgtgacca	10980
ccctgcgggt	gaactgcacg	tctaggggcg	agtagtccag	ggtttccctg	atgatgtcat	11040
acttatcctg	tcctcttttt	ttccacagct	cgcggttgag	gacaaactct	tcgcggtctt	11100
tcacgtactc	tttgatcgga	aaccgcgtcg	cctccgaagc	gtaagagcct	agcatgtaga	11160
actggttgac	ggccttgtag	gcgcagcate	ccttttctac	gggtagcgcg	tatgcctcgc	11220
cggccttcgc	gagcggaggt	tgggtgagcg	caaagggtgc	ccctaaccatg	actttgaggt	11280
actgtgattt	gaactcagtg	tcgtgcgcatc	cgccctgcctc	ccagagcaaa	aagtcctgtc	11340
gctcttttga	acgcgggttt	ggcaggggca	aggtgacatc	gttgaagagt	atctttcccg	11400
cgcgaggcat	aaagtgtcgt	gtgatcgagg	aggttccggg	caactcgga	cggttgttaa	11460
ttacctgggc	ggcgagcagc	atctcgtcaa	agcggctgat	gttgtgtgcc	acaaatgtaa	11520
gttccaaaga	gcgcgggagc	cccttgatgg	aaggccaatt	tttaagttcc	tcgtatgtga	11580
gctcttcagg	ggagctgagc	cgtgtctcgt	aaagggccca	gtctgcgaag	tgagggttgg	11640
aagcgcagaa	tgaactccac	aggttcaagg	ccattagcat	tgcaggttgg	tcgcgaaagg	11700
tcctaaactc	ggcgaactatg	ccattttttt	ctgggggtgat	gcagtagaag	gtacgcgggt	11760
cttgttccca	ggcgttcccat	ccaaggttccg	cggctaggttc	tcgcgcggcg	gtccactagag	11820
gtcatctcc	gcgcgaacttc	atgaccagca	tgaagggcac	gagctgcttc	ccaaagggcc	11880
ccatccaagt	ataggtctct	acatcgttag	tgaacaagag	acgctcgggt	cgaaggtacg	11940
agccgatcgg	gaagaaactcg	atctcccgcc	accagttgga	ggagtggtcg	tgtagtgtgt	12000
gaagtagaag	gtccctccga	cgggcgcgac	actcgtgctg	gcttttgtaa	aaacgtgcgc	12060

agtgactggca	ggcgtgcacg	ggctgtacat	cctgcacag	gttgacctga	cgaccggcca	12120
caagggaagca	gagtggaagt	ttgagccccc	cgctcggcgg	gtttggctgg	tggctcttcta	12180
cttcggctg	ttgtccttga	ccgtctgggt	gctcgagggg	agttacgggt	gatccgacca	12240
ccacgcgcg	cgagcccaaa	gtccagatgt	ccgcgcgcgg	cggtcgagag	ttgatgacaa	12300
catcgccgag	atgggagctg	tccatgggtc	ggagctcccg	cgcgctcagg	tcaggccggga	12360
gtctctcgag	gtttacctcg	catagccggg	tcaggcgccg	ggctaggctcc	aggtgatacc	12420
tgatttccag	gggtcgtgtg	gtggcgccgt	cgatggcttg	caagaggccg	catccccgc	12480
gcgcgactac	gggtacgcgc	ggcgggcggt	gggcccgcgg	gggtgctcctg	gatgatgcac	12540
ctaaaagcgg	tgacgcgggc	gggcccccg	aggtaggggg	ggctcgggac	ccgcgcggag	12600
agggggcagg	ggcacgctgg	cgccgcgcgc	gggcaaggag	tggtgctcgc	cgccgagggt	12660
gtcggcgaa	gcgacgacgc	ggcggttgat	ctcctgaatc	tgccgctctc	gcgtgaagac	12720
gcgggcgcgc	gtgagcttga	acctgaagaa	gagttcgaca	gaatcaattt	cggtgtcgtt	12780
gacgcgcgc	tgccgcaaaa	tctcctgcac	gtctcctgag	ttgtcttgat	agggcatctc	12840
ggccatgaac	tgctcgatct	cttctcctcg	gagatctccg	cgctcggctc	gctccacggt	12900
ggcgccgcgg	tcgttgagga	tgccggccat	gagctgcgag	aaggcggtga	ggcctccctc	12960
gttccagacg	cggtcgtaga	ccacgcccc	ttcgccatcg	cgggcgccga	tgaccacctg	13020
cgccgagatt	agctccacgt	gcggggcgaa	gacggcgtag	tttcgcaggc	gctgaagag	13080
gtagtgtagg	gtgtggcggt	tgtgttctcg	cacgaagaag	tacataaccc	agcgccgcaa	13140
ctgggttgat	ttgatctccc	ccaaggccct	aaggcgctcc	atggcctcgt	agaagtccac	13200
ggcgaggttg	aaaaactggg	agttgcgcgc	cgacaaggtt	aaactcctct	ccagaagacg	13260
gatgagctcg	gcgacagctg	cgccacctcc	gcgctcaagg	gctacagggg	ccctcttctc	13320
ttcttcaact	ttcttctcca	taaggccctc	cccttctctt	ttcttggcgc	gggtgggggg	13380
aggggggaca	cgccggcgac	gacggcgccg	cgggagggcg	tcgacaagag	gctcgatcat	13440
ctcccccg	cgagcgccga	tggtctcgtg	gacggcgccg	ccgttctcgc	ggggggcgag	13500
ttggaaagac	cgcccgctca	tgctccgggt	atgggttgct	ggggggcgct	cgctcgccag	13560
ggatacgccg	ctaacgatcg	atctcaacaa	ttgtttgtga	ggtaactccg	caccgaggga	13620
ctcgagcgag	tccgcctcga	ccggatccga	aaacctctcg	agaaaggcgt	ctaaccagtc	13680
acagtgcgaa	ggtaggctga	gcaccgcgcg	ggcgccgcgc	ggcgccgcgt	cggggttgtt	13740
tctggccgag	gtcctgctga	tgatgtaatt	aaagttagcg	ctcttgagac	ggcgcatggt	13800
cgacgaagac	accatgtctc	tggttcgcgc	ctgctgaatg	cgacggcggt	cgcccatgcc	13860
ccaggcttcg	ttttgacatc	ggcgacggct	tttgtagtag	ctttgcatga	gcctttctac	13920
cgccactctt	ttcttctcct	ctcttcttcc	tgcatctctt	gcattctatc	ctcgccggcg	13980
ggcgaggttt	ggccgtaggt	ggcccccctt	tcctcccatg	cggtgaccc	cgaaagccct	14040
catcgccgta	agcagggcca	ggtcgccgac	aaagcgctcg	gctaattatg	ccctgctgac	14100
ctcgctgaag	gtagactgga	agtcgtccat	gtccacagaa	cggtggtatg	ggcccggtgt	14160
gatggtgtaa	gtgcagttgg	ccataacgga	ccagtttaac	gtctgtgagc	ccggttgagc	14220
gagctcgggt	tacctgagac	gcgagtaagc	ctttgagtcg	aagacgtagt	aggtgcaagt	14280
ccgcaccagg	tactgtgtac	ccaccaaaaa	gtggcgccgc	ggctggcggt	agagggggcca	14340
gcgtagggtg	ggccgggctc	ggggggcgag	gtcttcccaac	ataaaggcgt	gatatccgta	14400
gatgtccctg	gcacatccag	tgatgcgcgc	ggcggttggt	agggcgccgc	gaagatcacg	14460
gacgcggttc	cagatgttgc	gcagcgccga	aaagtgtctc	atggtcgggg	cgctctggcc	14520
ggtcaggcgc	ggcgagtcgt	tgacgctcta	gacgcgtgca	aaggagagcc	tgtaacggg	14580
caactctccg	tggtcttggt	gataaattcg	caagggtatc	atggcgagag	accgggggtt	14640
gaaccccgga	tcggccgcgt	cgccgtgac	catcggttta	ccgcccgctg	gtcgaaacca	14700
ggtgtgcgac	gtccagacaac	gggggagcgc	tccttttggc	ttccttccag	ggcgccggga	14760
tgctcgctga	gtctttttgg	ccactggccg	cgccggcggt	aagcggttag	cgctgaaagc	14820
gaagacatta	agtggctcgc	tcctctgtag	cgaggggtta	ttttccaaag	gttgagtcgc	14880
gggacccccg	gttcgagctc	cgggcccgcc	ggactgcggc	gaacgggggt	gtctctcccc	14940
gtcactgcag	accccgcttg	caaatctctc	cggaacacag	gacgagcccc	ttttttgtct	15000
ttccagagtg	ctcccggtgc	tgccgcagat	gcgcgccctc	ctccagcagc	ggcaagggca	15060
agagcagcgg	cagacatgca	gggcacccct	cccttctctc	accgctctag	gagggggcaac	15120
atccgcgctc	gacgcggcgg	cagatgggtg	ttacgaaccc	ccgcggcgcc	ggacccggca	15180
ctactggagc	ttggagagag	gcgagggcct	ggcgccgcta	ggagccgcct	ctccgtgagc	15240
acacaaagg	gtcagctgga	agcgtgacac	gcgcgagggc	tacgtgccgc	ggcagacctc	15300
gttttcgcgac	cgcgaggagg	aggagccgga	ggagatgcgg	gatcgaaagt	ttcatgcagc	15360

gcgcgagttg	cggcatggcc	tgaacccgga	gcggttgctg	cgcgaggagg	accttgagcc	15420
gcagccgcgg	accgggatta	gtcccgcgcg	cgacacacgtg	gcggccgcgc	acctggtaac	15480
gcgcatacag	cagacgcgtga	accaggagat	taactttcaa	aaaagcttta	acaaccacgt	15540
gcagcagcgt	gtggcgcgcg	aggagggtgc	tataggactg	atgcattctgt	ggagctttgt	15600
aagccgcgctg	gagcaaaaacc	caaatagcaa	gcgcgtcatg	gcgcagctgt	tccttatagt	15660
ggagcacagc	agggaacaacg	aggcatcag	ggatgcgctg	ctaaacatag	tagagcccgga	15720
cggccacagc	ctgctgcgatt	tgtataaact	tctgcagcag	atagttggctg	aggagccgag	15780
cttgagcctg	gctgacaagg	tggccgccat	taactattcc	atgtcagtc	tgggcaagt	15840
ttacgcgcgc	aagatataacc	atacccttca	cgttccccata	gacaaggagg	taaaagatcga	15900
ggggttctctg	atgcgcgatgg	cgtgcaaggt	gcttaacctg	agcgacgacc	tgggcgttta	15960
tcgcacaagag	cgcattccaca	aggccgtgag	cgtgagccgg	cgccgcggagc	tcagcgaccgc	16020
cgagctgatg	cacagccctgc	aaaggccctt	ggctggccagc	ggcagcgccg	ataagagagcc	16080
cgagctctac	tttgagcgcg	gcgctgacct	gcgctggggc	ccaagccgac	gcgccttsga	16140
ggcagctggg	gcgcgacctg	ggctggcggt	ggcaccgcgc	cgccgtggca	acgtctggggc	16200
cgtggaggaa	tatgacgagg	acgatgagta	cgagccagag	gacggcgagt	actaagcggg	16260
gatgtttctg	atcagatgat	gcaagacgca	acggaccggc	cggtgcgggc	ggcgctgcag	16320
agccacgcgt	ccggccttaa	ctccacggac	gactggcgcc	aggtcatgga	cgcctcatgt	16380
tcgctgactg	cgcgcacccc	tgcagcgttc	cgccagcagc	cgacggccaa	cggcctcttc	16440
gcaacttctgg	aaagcgggtgt	cccgcgcgcc	gcaaacccga	cgacagagaa	ggctgctggc	16500
atcgttaaacg	cttcggccgga	aaacaggggc	atccgcggcg	atgagggccg	cctgggtctac	16560
gacgcgctgc	ttcagcgcgt	ggctcgttac	aaacagcaga	acgtgcaagc	caactctgac	16620
cgccctgggtg	gggatgtgtgc	cgaggccgtg	gcgcagcgtg	agcgcgcgca	gcagcagggc	16680
aaactgggctc	ccatgggttgc	actaaacgcc	ttcctgagta	ccagcgcgcg	caactgtgcc	16740
cggggacagc	aggactacac	caacttttgt	agcgcattcg	ggcttaattgt	gactgagaca	16800
ccgcaaaagtg	agggttatca	gtccggggcca	gactattttt	tcagacaccg	tagacaaggc	16860
ctgcagacgc	taaacctlag	caggccttct	aagaaactgc	aggggctgtg	gggggtgcgc	16920
gctccacagc	gcgacccgcg	gaccgtgtct	agcttgctga	cgcccaactc	gcgcctgttg	16980
ctgctgtcaa	tagcccccct	caccgacagt	ggcagcgtgt	cccgggacac	atacctaggt	17040
cacttgctga	cactgtaccg	cgaggccata	ggtcaggcgc	atgtggacga	gcatactttc	17100
caggagatta	caagtgttag	ccgcgcgctg	gggcaggagg	acacgggcag	cctggagcca	17160
accttgaaact	acctcgtgac	caaccggcgc	caaaaaatcc	cctcgttgca	cagtttaaac	17220
agcgaggagg	agcgcatttt	gcgctatgtg	cagcagagcg	ttagccttaa	cctgatgcgc	17280
gcgggggttaa	cgcccagcgt	ggcgtgggac	atgacccgcg	gcaacatgga	accggccgatg	17340
tatgcctcaa	accggccggt	tatcatatgc	ctaattggact	acttgcatcg	cgccggccgcc	17400
gtgaaccccg	agtattttcac	caatgccatc	ttgaacccgc	actggctacc	gcgccctggg	17460
ttctacacgc	ggggattcga	ggtcgccgag	ggtaacgatg	gatttcctctg	ggacgacata	17520
gagacagcgc	tgtttttccc	gcaaccgcag	acctcgtag	agttgcaaca	acgcgagcag	17580
gcagaggcgc	cgtgcgga	ggaagagctc	cgacggccaa	cgagcttgct	cgatctaggc	17640
gctcgccgcc	gtcgttcaga	tgtctatgac	ccatttccaa	gcttgatagg	gtctcttacc	17700
agcctatcgca	ccaccgcgcc	gcgcctgctg	ggcgaggagg	agtaacctaa	caactcgtcg	17760
ctcagcagca	agcgcgaaaa	gaacctgcct	ccgcgcttcc	ccaacaacgc	gatagagagc	17820
ctagtggaaca	agatagtag	atggaagacg	tatgcgcagg	agcacaggga	gtgcccgcgc	17880
ccgcgcgcgc	accacgcctg	tcaaaaggcag	gacgctcagc	ggggtctggg	gtggaggagac	17940
gatgactcgc	cagcagcagc	cagcgtcttg	gatttgggag	ggagtggcaa	cccggtttgca	18000
caaccttcgc	ccagctgggg	gaggaatgtt	taaaaaaaag	catgatgaa	aataaaaaaa	18060
tcaccacaagc	cattgctggg	agcgttggtt	ttctgttatt	ccctctagta	tgcggcgcg	18120
ggcgatgat	gagggaaggt	ctcctcccct	ctacgagagc	gtggtgagcg	cgccgcaggt	18180
ggcgcgcgcg	ctgggtttcac	ccttcgatgc	tcctcctggc	ccgcgcttcg	tgcctccgcg	18240
gtacctcggc	cctacccggg	ggagaaaacg	catccgtaac	tctgatgtgg	acacctctat	18300
cgacaccacc	gtgtgttacc	tttgtggaca	caagtccaac	gatgtggcat	ccttggaacta	18360
ccagaacgac	cacagcaact	ttctaaacac	ggtcatatcaa	aaacatgact	acagcccggg	18420
ggagggcaagc	acacagacca	tcaactttga	gcagcgtgcg	ccctggggcg	gcgacctgaa	18480
aaccatcctg	catcacaaca	tgcctaaatg	gaacgaggtc	atgtttacca	ataagtttaa	18540
atggcggggtg	atggctgcgc	gcgctgttac	taaggacaaa	caggtggagc	tgaatacga	18600
gtgggtggag	ttcacgctgc	ccgagggcga	ctactccgag	accatgacca	tagaccttat	18660

gaacaacgcg	atcgtggagc	actacttgaa	agtgggcagg	cagaacgggg	ttctggaagg	18720
cgacatcggg	gtaaaagt	atccccgcaa	cttcagactg	gggtttgacc	cagtcactgg	18780
ctttgtcatg	cctgggggat	atacaaacga	agccttccat	ccagacatca	ttttgtcgcc	18840
aggaatgggg	gtggactcca	cccacagcgc	cttgagcaac	ttgtttggga	tcgcgaagcg	18900
gcacacccttc	caggagggct	ttaggatcac	ctacagatgac	ctggagggtg	gtaacatctc	18960
cgcaactgttg	gatgtggagc	cttcaccaggc	aagcttgaaa	gatgacacgc	aacaggggcg	19020
gggtggcgca	ggcggcgcca	acaacagctg	cagcggcgcg	gaagaacact	ccaacgcggc	19080
agctggcgca	atgcagccgg	tggaaggacat	gaacgatcat	gccattcgcg	gcgacacctt	19140
tgccacacgg	ggcgaggaga	agcgcgctga	ggcggaggca	ggcgccgaag	ctgcgcgccc	19200
cgcctggggg	gctgcacaa	ccgagggtga	gaagcctcag	aagaaacccg	tgattaaacc	19260
cccgacagag	gacagcaaga	aacgcagtta	caacctaa	agcaatgaca	gcaccttcac	19320
ccagtcacgc	agctgggtacc	ttgcatacaa	ctacggcgac	cctcaggccg	ggatccgcct	19380
atggaccctg	ctttgcactc	ctgcagtaac	ctgcggctcg	gagcaggtat	actggctcgt	19440
gcccgcacatg	atgcaagcag	ccgtgacctt	ccgctccacg	cgccagatca	gcaactttcc	19500
gggtgggtggc	cgcgagctgt	tgcccggtga	ctccaagagc	ttctacaacg	accaggccgt	19560
ctactcccaag	ctcatccgcg	agtttacctc	cttgaccacc	gtgttcaatc	gctttcccgca	19620
gaaccagatt	ttggcgcgcc	cgcgcagccc	cacctacacc	accgtcagtg	aaaaagcttc	19680
tgctctcaca	gatcacggga	cgtaccgcgt	gcgcaacagc	atcggaaggag	tcgacgcagt	19740
gacacttact	gagccagagc	gcgcgacctg	cccttcagtt	tacnaagccc	tgccgatagt	19800
ctgcgcgcgc	gtcctatcga	gcccgcactt	ttgagcaagc	atgtccatcc	ttatatcgcc	19860
cagcaataaac	acaggctggg	gcctgcgctt	cccaagcaat	atggttggcg	ggcgcaagaa	19920
gcgcctccgac	caacaccag	tgccgctgctg	cgggcactac	cgccgcgcct	ggggcgcgca	19980
caacgcgcgc	cgcactggcg	gcaccacgct	cgatgacgcg	ctcgacgcgg	tggtggaggga	20040
gcgcgcgaac	tacacgccca	cgccgcgcgc	agtgctccacc	gtggacgcgg	ccatttcagac	20100
cgtgtgtgcg	ggagcccgcg	gctacgctaa	aatgaagaga	cgccggaagg	cgctagcagc	20160
tgcccaaccgc	cgcgcacccg	gcactgcgcg	ccaacgcgcg	ggcgcgccgc	tgcttaaccg	20220
ctgcactcgc	accgcgcgac	ggcgcgccat	gcgaagccgt	cgaaggctgg	cgccgggtat	20280
tgctactgtg	ccccccaggt	ccaggcgacg	agcggccgcg	gcgacgcgcg	cgccattagg	20340
tgctatgact	caagggtcgca	ggggcaacgt	gtactgggtg	cgcgactcgg	ttagcgggct	20400
cgcggtgccc	gtgcgcaccc	gccccccgcg	caactagatt	gcaataaaaa	actacttaga	20460
ctcgtactgt	tgatgtatc	cagcggcgcg	ggcgcgcctc	gaagctatgt	ccaagcgcaa	20520
aatcaaaagaa	gagatgtctc	aggtcatcgc	gccggagatc	tatggccccc	cgaagaaggga	20580
agagcaggat	tcaagccccc	gaaagctaaa	cggggtcaaa	aagaaaaaga	aagatgatga	20640
tgatgatgaa	cttgacgacg	aggtggaact	gttgcacgcg	accgcgcaca	ggcgacgggt	20700
acagtgaggaa	ggtcgacgcg	taagacgtgt	tttgcgaccc	ggcaccacg	tagtctttac	20760
gcccggtgag	cgtctacccc	gcacctacaa	gcgcgtgtat	gatgaggtgt	accgcgacga	20820
ggacactgctt	gagcagggca	acgagcgccg	cggggagtgt	gcttacggaa	agcggcataa	20880
ggacatctgt	gctgtgccgc	tggaagcagg	caaccacaac	ctcagcttaa	agcccgtagc	20940
actgcagcag	ggctgtcccg	cgcctgcacc	gtccgaagaa	aagcgcgcgc	taaacgcgca	21000
gtctgtgagc	ttggcaccca	ccgtgcagct	gatggtaacc	aagcgtcagc	gactggaaag	21060
tgctctggaa	aaaatgaccg	tggaagctgg	gctggaagcc	gaggtccgcg	tgccggcaat	21120
caagcagggtg	gcaccggggc	tgggcggtga	gacccgtgag	gttcagatca	ccaccaccag	21180
tagcactagt	atggccactg	ccacagaggc	catggagaca	caaacgtccc	cggttgcttc	21240
ggcggtggca	gatgcgcggg	tgacggcgcg	cgcctgcggc	gcgtccaaga	cctctacgga	21300
gggtgcacaa	gaccgcgtga	tglttctgtg	ttcagccccc	cggcgctccg	gcgcttcaag	21360
gaagtaccgc	gcgcgcagcg	cgtactcgcc	cgaatatgcc	ctacatctct	ccatcgccgc	21420
tacccccggc	tatcgtggct	acacctaccg	ccccagaaga	cgagcaacta	cccgacggcg	21480
aaccaccact	gcgcgcgcgc	gcgcgcgcgc	ccgtgcgcgc	cccgctgctg	ccccgatttc	21540
cgtgcgcagc	gtggctcgcg	agggaaggag	gacctcggtg	ctgcgaacag	cgccgttaca	21600
ccccagcttc	gtttaaaagc	cggctcttgt	ggttcttgca	gatattggcc	tcacctgctg	21660
ctctcgtttc	cgggtgcggg	gattccgcgg	aagaatgcac	cgtaggaggg	gcattggcgg	21720
ccaagccctg	gcggcgcgca	tgctcgtcgc	gcaccaccg	cgccgcgcgc	cgctgcaccg	21780
tcgatagctg	ggcgggtatc	tgccctccct	tattccactg	atcgccgcgc	cgaattggcg	21840
cgtgcggcga	gtgtctccg	tgcccttgca	ggcgacagga	cactgattaa	aaacaagtta	21900
catgtggaaa	aatcaaaaa	aaagtctgga	ctctcacgct	cgccttgcttc	tgtaactatt	21960

ttgtagaatt	gaagacatca	acttttgcgtc	actggcccg	cgacacggct	cgcccccgtt	22020
catgggaaac	tggcaagata	tcggcaccag	caatatgagc	ggtggccctt	tcagctgggg	22080
ctcgctgttg	agccgatta	aaaatttcgg	ttccgcgctt	aaagactatg	gcagcaaaag	22140
ctggacacgc	agcacaggcc	agatgctgag	ggacaagtgt	aaagagcaaa	atttccaaca	22200
aaagttggta	gatggcgttg	ctcttggtcat	tagcgggggt	gtggacctgt	ccaaccagcg	22260
agtgcaaaat	aaatttaaca	gtaagcttga	tcgccgccct	cccgtagagg	agcctccacc	22320
ggccgtggag	acagttgttc	cagagggggt	tggcgaaaag	cgctccgcag	ccgcacagga	22380
agaaaactct	gtgacgcaaa	tagacgagcc	tcctctgcat	gaggagggac	taaaagcagg	22440
cttgcccacc	accggtccca	tcgctgccat	ggtacccgga	gtgctggggc	agccacacac	22500
ctctaaagct	gaacttgccc	ccccggccga	caccacagag	aaacctgtgc	tgccagggcc	22560
gtccgcgctg	gttgttaacc	gtcctagccg	cgctgcccgt	cgccgcgcgt	ccagcggtcc	22620
cgatctgttg	cgcccgctag	ccagtgccaa	ctggcaaaag	acactgaaca	gcactgtggg	22680
tttgggggtg	caatccctga	agccgcgacg	atgcttctga	tagctaacgt	gtcgtatgtg	22740
gtctcatgat	gcgtccatgt	cgccgcgaga	ggagctgctg	agccgcgcgt	cgccgccttt	22800
ccaagatggc	taccctctgc	atgatgccgc	agtgtcttta	catgcacatc	tcggggccag	22860
accgctcgga	gtacctgagc	ccgggctggg	tgcagtctgc	ccgcgccacc	gagagctact	22920
tcagctcgaa	taacaagttt	agaaacccca	cggtggcgcc	tacgcacagc	tacgaccagag	22980
accggtctca	gggtttgacg	ctgcgggtta	tcgccgtgga	ccgcgaggat	actgcgtact	23040
cgtaacaagg	cggtgttaacc	ctagctgttg	gtgataacgc	tgtctgtagac	atgggtctca	23100
cgactcttga	catccgcggc	gtgctggaca	ggggccctact	ttttaagccc	tactctggca	23160
ctgcctacaa	cgcatctggc	cccaagggtg	ccccaactc	gtgcgagtgg	gaacaaaatg	23220
aaactgcaca	agtgtgatgt	caagaacttg	acgaagagga	gaatgaagcc	aatgaagctc	23280
agggcggaga	acaggaacaa	gctaaagaaa	cccatgtata	tgccagggct	ccactgtccg	23340
gaataaaaa	aactaaagaa	ggtctacaaa	taggaactgc	cgacgcocaa	gtagcaggtg	23400
ccggcaaaag	aattttcgca	gacaaaactt	ttcaacctga	accacaagta	ggagaactct	23460
aatggaacga	agcggatgcc	acagcagcgt	gtgggaaggt	tcctaaaaag	acaaactcca	23520
tgaaacctcg	ctatggtctca	tacgctagac	ccacaatttc	caacggcgga	caggggcgta	23580
tggttgaaac	aaatggtaaa	ttggaaagtc	aagtcgaaat	gcaatttttt	tcacatccca	23640
caaatgccac	aaatgaagtt	aacaataatc	aaaccaacgt	tgtattgtac	agcggaagatg	23700
taaacactgga	aactccagat	actcatcttt	cttataaacc	taaaatgggg	gataaaaaatg	23760
ccaaagtcac	gcttggacaa	caagcaatgc	caaacagacc	aaattacatt	gctttttagag	23820
acaattttat	tggtctcatg	tattacaaca	gcacaggtaa	catgggtgtc	cttgtctgttc	23880
agggcatcgca	gttgacagct	gtttgagatt	tgcaagacag	aaacacagag	ctgtcctacc	23940
agcttttgtc	tgattcaatt	ggcgacagaa	caagataactt	ttcaatgtgg	aatcaagctg	24000
ttgacagcta	tgatccagat	gtcagaatta	ttgagaaaca	tggaactgag	gatgagttgc	24060
caaatatttg	ctttctctctt	ggtggaattg	ggattactga	cacttttcaa	gctgttaaaa	24120
caactcgtgc	taacgggggac	caaggcaata	ctacctggca	aaaagattca	actatttgacg	24180
aacgcgaatga	aatagggggtg	ggaataaact	ttgccatgga	aattaacctg	aatgccaacc	24240
tatggagaaa	ttctctttac	tcataattg	cgctgtacct	ccagacaaag	ctaaaataca	24300
acccaccaca	tggtggaata	tcgacaaac	ccaacacctc	cgactacatg	aaacaagcag	24360
tggtgctctc	tggtctgtga	gactgctaca	ttaaccttgg	ggcgcgctgt	tcctctggact	24420
acgtggacaa	cgttaatccc	tttaaccacc	accgcaatgc	gggcctgcgt	taccgctcca	24480
tggtttgtgg	aaaacgcgcg	tacgtgccct	ttcacattca	ggcgccccaa	aagttttttg	24540
ccattaaaaa	ctctctctct	ctgcacagct	catcacacata	tgaatggaaac	ttcagggaag	24600
atgtttaaac	gggtctgcag	agctctctgg	gaacagcaat	tagagttgac	ggggctagca	24660
ttaaggttga	cagactttgt	cttttaagcca	ctttcttccc	catggccacc	aaacagggct	24720
ccaagctgga	agccatgctc	agaaatgaca	ccaacggacc	gtcctttaat	gactaccttt	24780
ccgcgcgcaa	catgctatat	cccataccgc	ccaacggacc	caacgtgccc	atctccatcc	24840
catcgcgcaa	ctggggcagca	ttctcggttt	gggccttacc	acgtctgaag	acaaagggaa	24900
ccctctctct	gggatacagg	taagacacct	atacacatac	ctctggctcc	ataccatacc	24960
ttgacgggac	ctctctabctt	aatcacacct	ttagaaggtg	ggccattact	tttgactctt	25020
ctgtttagctg	cgccggcgaac	gaccgctgc	ttactcccaa	tgaagttgac	cttaagcgct	25080
cagttgacgg	ggagggctat	aacytagctc	agtgcaacat	gacaaaggag	tggttctcatg	25140
tgcagatgtt	ggcccaactac	aatattggct	accagggctt	ctacattcca	gaagactaca	25200
aagaccgcat	gtactcgttc	ttcagaanaat	tcacggccat	gagccggcaa	gtgttgagag	25260

atactaaata	caaaagattat	cagcagggttg	gaattatcca	ccagcataac	aactcaggct	25320
tcgtaggctta	cctcgtctccc	acatgcgcgc	agggaacaagc	ttaccccgct	aatgttccct	25380
accactaat	aggcaaaaac	gggttgata	gtattatocca	gaaaaagttt	ccttgccacc	25440
gcacctgtg	gcgcataccc	tttcccaagta	actttaatgtc	catgggtgag	ctccacagacc	25500
tgggccaaaa	cctctctctac	gcaaaactccg	cccacgcgc	agacatgacc	tttgagggtg	25560
atcccatgga	cgagccccacc	cttcttttatg	ttttgtttga	agtcttgcac	gttggtccgtg	25620
tgccacagacc	gcaccgcggc	gtcatcgaga	ccgtgtacat	ggccacgcgc	ttctcgccgc	25680
gcacgccacc	aacataaaga	agcaagcaac	atcaacaaca	gctgccgcga	tgggctccag	25740
tgagcaggaa	ctgaaaagcca	ttgtcaaaaga	tcttggttgt	gggccatatt	ttttgggcac	25800
ctatgacaga	cgcttcccag	gctttgtttc	cccacacaag	ctcgccctgc	ccatagttaa	25860
caaggccggt	cgagagactg	ggggcgatca	ctggatggcc	tttgccggga	accgcgcctc	25920
aaaaacatgc	tacotctttg	agccctttgg	ctttcttgac	caacgtctca	agcaggttta	25980
cgagttgag	tacagtgac	tctcgccgc	tagcgccatt	gcctctcccc	accgcgcgtg	26040
tataacgctg	gaaaagttca	cccaaaagct	gcagggggccc	aactcgccgc	ctgtgggctc	26100
atctcgtgc	atgtttctcc	acgcctttgc	caactggccc	caaaactcca	tgatcacaa	26160
cccaccatgc	aaccttatta	ccgggggtacc	caactccatg	cttaacagtc	cccaggtaca	26220
gcccaccctg	cgccgcacac	aggaacagct	ctacagcttc	ctggagcgcc	actgccctta	26280
cttcgcagc	cacagtgcgc	aaattaggag	cgccacttct	ttttgtcact	tgaaaaacat	26340
gtaaaaataa	tgtaactagga	gacactttca	ataaaggcaa	atgtttttat	ttgtacactc	26400
tcgggtgatt	attttacccc	accctttgccc	tctgcgcgt	ttaaaaatca	aagggtgtct	26460
gcgcgcctac	gctatgcgcc	actggcaggg	acacgtttgc	atactggtgt	cttagctctc	26520
acttaaaact	aggacacaac	atccgcggca	gctcggtgaa	gtttttcactc	cacaggtgcg	26580
gcactatcac	caacgcgttt	agcaggttcg	gcgcgatata	cttgaagtgc	cagtgtgggc	26640
ctcgcgcctg	cgccgcgcag	ttgcgatata	cagggttaca	gcactggaaac	actatcagcg	26700
ccgggttggtg	cacgcctggcc	agcacgctct	tgctggagat	cgatccgcgc	tcagagttct	26760
ccgcgctgct	cagggcgaaac	ggagtcaact	ttggtagctg	cttcccccac	aagggttgcat	26820
gcccaggctt	tgagttgcac	tcgcaccgta	gtggcatcac	aaggtgaccg	ttgccagtct	26880
ggggcgttag	atcacagccg	tgcatgaaag	cttgatctct	cttaaaagcc	acctgagcct	26940
ttgcgcttc	agagaagaaac	atgcgcgaag	acttgccgga	aaactgatgt	gcgggacagg	27000
ccgcgtctatg	cacgcagcac	cttgccgtcgg	tggtggagat	ctgcaccaca	tttcggcccc	27060
accggttctt	cacgatcttg	gccttgctag	actgctcctt	cagcgccgc	tgcccgctttt	27120
gcctgctcac	atccatttca	atcacgtgct	ccttatttat	cataatgctc	ccgtgtagac	27180
acttaaggct	gccttcgata	tcagcgccgc	ggcgacgcca	caacgcgcag	cccggtggct	27240
cggtgtgctt	gtaggttaac	tctgcaaaag	actgcaggtta	cgccctcagg	aatgcgccca	27300
tcactgcac	aaagggtctg	ttgttggtga	aggttcagctg	caaccgcgcg	tgctcctcgt	27360
ttgcgcaggt	cttgcatagc	gcgcgcagag	cttccactgt	gtccagcagt	agcttgaagt	27420
ttgcctttag	atcgctatcc	acgtggtaact	tgctcatcaa	ccgcgcgcga	gcctccatgc	27480
ccttctccca	cgacagacac	atccggagcc	tcagcggggt	tatcacccgt	ctttcacttt	27540
ccgcttctac	ggactttctc	ttttctctct	gcctccgat	accgccacct	actgggtcgt	27600
cttcaattcac	cgccgcgacc	tgctgcgtac	ctcccttgcc	gtgcttgatt	agcaccggtg	27660
gggttgctgaa	accaccacct	tgtagcgcca	catcttctct	ttcttctcgt	ctgtccacga	27720
tcactctgg	ggatggcggg	cgctcgggct	tggaagaggg	gcgcttcttt	ttctttttgg	27780
agcaaatggc	caaatccccc	gtcgaggtcg	atggccgcgc	gctgggtgtg	cgcggcacca	27840
gcgcactctg	tgacagagtc	ttctcgtctc	cggaactgag	accgcgctc	cagctgcttt	27900
ttggggggcg	gcgggggaggc	ggcggcgacg	gcgacgggga	cgagacgctc	tcctaggttg	27960
gtggagctgc	cgccgcaccc	cgctccgcct	cggggggtgt	ttcgccctgc	tcctctctcc	28020
gactgcccac	ttctctctcc	tataggcaga	aaaagatcat	ggagtccagc	gagaaaggag	28080
acagcgcctaac	gcgccctctt	gagttcgcca	ccaccgcctc	accgatgccc	gcgaagcgcc	28140
ctaccacctt	ccccctcgag	gcaccccgcc	ttgaggagga	ggaagtgtat	atcgagcagg	28200
accaggtgtt	tgtaagcgaa	gacgacgaag	atcgctcagt	accaacagag	gataaaaagc	28260
aaagccagga	cgagcgcagag	gcaaacgagg	aacaagtctg	gcggggggag	caaaagcatg	28320
gcgactacct	agagtgtggga	gacgacgtgc	tgttgaagca	cttcgacgcg	ctagtcgcca	28380
ttatctgcga	cgctgttgcaa	gacgacgagc	atgtgccccct	cgccatagcg	gagtgcagcc	28440
ttgcctacga	acgcgcacctg	ttctccaccg	gcgtaccccc	caaacgccaa	gaaacgcgca	28500
catgcgagcc	caaccgcgcg	ctcaactctt	acccttgatt	tgcgctgcca	gaggtgcttg	28560

ccaccatcatca	cattctttttc	caaaactgca	agataccccc	atccctgccgt	gocaaaccgca	28620
ccgcagcgga	caagcagctg	gccttgcggc	agggcgctgt	catacctgat	atcgccctgc	28680
tcgcaggaagt	gcgcaaaatc	tttgagggtc	ttggacgcga	cgagaagcgc	gcggcaaacg	28740
ctctgcaca	agaaacacgc	gaaaatgaaa	gtocacttgg	agtgctgggt	gaacttgagg	28800
gtgacacgc	gcgcctagcc	gtgctgaaac	gcagcatcga	ggtaaccacc	tttgctctacc	28860
cggcaacttaa	ctaccccccc	aaagttatga	gcacagtcac	gagcgagctg	atcgtgcgcc	28920
gtgcacgacc	cctggsgagg	gatgcacact	tgcagaacac	aaccgagggg	ggcctaccgc	28980
cagttggcgca	tgagcagctg	gcgcgctggc	ttgagacgcg	cgagcctggc	gaectggagg	29040
agcgacgcga	ctaatgagt	gcgcgagtgc	ttgttaccgt	ggagctttag	tgcacgcagc	29100
ggttcttttgc	tgaccgggag	atgcagcgca	agctagagga	aacgttgacac	tacacacttcc	29160
gccagggtcta	cgtgcgccag	gcctgcacaaa	tttccacacgt	ggagctctgc	aacctggtct	29220
ctcactcttgg	aaattttgcac	gaaaaaccgc	ttgggcaaaa	cgtgctctcat	tccacgctca	29280
agggcgaggcg	gcgcgcgcgac	tacgtccgcg	actcggttta	cttattttctg	tgctacacct	29340
ggcaaacggcg	catggcgctg	ttggcagcagt	gcctggagga	gcgcacacctg	aaggagctgc	29400
agaagctgct	aaagcaaaac	ttgaaaggacc	tatggacggc	cttcaacgag	cgctccgtgg	29460
cgcgcgacct	ggcgagcatt	atcttccccg	aacgcctgct	taaaacccctg	caacagggctc	29520
tgccgaactt	ccacagtcac	agcatgttgc	aaaactttag	gaactttatc	ctagagcgctt	29580
cagggaattct	gcggccgacc	tgtctgtcgc	ttcctagcga	ctttgtgccc	atataagtacc	29640
gtgaatgccc	ctccgcgctt	ttgggttact	gctacctctc	gcagctagcc	ctgaactctgt	29700
ctcaccactc	cgacatcatg	gaagacgtga	cggttgacgg	ctcactggag	ttgtcaactgtc	29760
gctgcacact	atgcaccccg	caaccgtccc	ttggtctgcaa	ttcacacactg	cttagcgaaa	29820
gtcaaatat	cgttactctt	gagctgcagg	gtccctcgcc	tgacgaaaaa	ttcgcggctc	29880
cggggttgaa	actcactccg	gggctgtgga	cgtcggttta	ctcttcgaaa	tttgtactct	29940
aggactacca	cgccacgag	attaggttct	acgaagacca	atcccccccg	ccaaaatggcg	30000
agctttaccgc	ctcgtctcat	accacgggccc	acatctcttg	ccaattgcaa	gcattataca	30060
aagcccgcca	agagtttctg	ctacgaaaag	gacggggggg	ttacttggac	ccccagttcg	30120
cgagggagct	caacccaatc	ccccgcgccg	cgcagcccta	tcagcagccg	cgggcccttg	30180
cttccacgga	tggcacccaa	aaagaagctg	cagctgcccg	cgccgcaccc	caaggacgag	30240
gaggaatact	gggacagtca	ggcagaggag	gtttttggacg	agggaggaga	gatgatggaa	30300
gactgggagca	gcttagacga	ggaagcttcc	gaggccgag	aggtgtcaga	cgaaacacgc	30360
tcaccctcgg	tcgcatctcc	ctcgccggcg	ccccagaaat	cgccaacctg	tcccagcatt	30420
gctacaacct	cgctctctca	ggcgccgcgc	gcactgcccg	ttcgccgacc	caaccgtaga	30480
tgggacacca	ctggaaccag	ggccggttaag	tctaaagcgc	cgcccgctgt	agcccaagag	30540
caacacacgc	gccaaggcta	cgctctgtgg	cggttgcaac	agaaagccat	agttgtcttc	30600
ttgcaagact	gtggggggcaa	catctctctc	gcccgcgct	ttcttctcta	ccatcacggc	30660
ttggcccttcc	ccgttaacat	cctgcattac	taccgtctac	ttcacagccc	ctactgcacc	30720
ggcgagcagc	gcgacaacag	cagcgccac	gcagaagcaa	agggcaccgg	atagcaagac	30780
cttgcacaag	cccaagaaat	ccacagcgcc	ggcagcagca	ggaggaggag	cactgcgctc	30840
ggcgcccaac	gaaccctgat	cgaccgcgca	gcttagaaac	aggaattttc	ccactctgta	30900
tgctatattt	caacagagca	ggggccaaga	acaagagctg	aaaataaaaa	acaggtctct	30960
gcgctccctc	accgcgagct	gcctgtatca	caaaagcgaa	gatcagcttc	ggcgacgct	31020
ggaagcggcg	gaggtctctc	tcagcaaaata	ctgcgcgctg	actcttaagg	actagttctg	31080
gcgcctctct	caaatttaag	cgcgaaaact	acgtctatct	cagcgccacc	accggcgccc	31140
agcacctgtc	gtcagcgcca	ttatgagcaa	ggaatttccc	acgccttaca	ttgtggagtta	31200
cccgaccaca	atgggacttg	cggctggagc	ttcccaagac	tactcaaccc	gaataaacta	31260
catgagcgcg	ggagccacca	tgatatggcg	ggtcaacgga	atccgcgccc	acogaaacgc	31320
aaattctctc	gaacaggcg	ctattaccac	caacactctg	gaataaccta	atcccgtag	31380
ttggcccgct	gccttgggtg	accaggaaa	tcccgctccc	accactgtgg	tacttccag	31440
agagcgccag	gcgaaggttc	agatgactaa	ctcaggggcg	cagcttggcg	gcggctttcg	31500
tcacagggtg	cggctgcgcc	ggcagggtat	aaactcacgt	aaactcagag	ggcgaggatg	31560
tcagctcaac	gacgagtcgg	tgagctctct	ttttgttctc	cgtccggagc	ggacacttca	31620
gatcgccggc	gctggcgctg	cttcatltac	gcccctgac	ggcatctcaa	ctctgcagac	31680
ctcgtctctg	gagcgcgctc	cgggaggaat	ttgaactcta	caattttattg	aggaagttgt	31740
gcctctgggt	tacttcaacc	ccctttcttg	acctcccggc	catcaccccg	accagtttat	31800
ttccacactt	gacgcggtaa	aaagctcgcc	ggcaggtac	gactgaatga	ccagtgagga	31860

ggcagagcaa	ctgcgcctga	cacacctcga	ccactgcgcg	gcgccacaagt	gctttgcccc	31920
cggtcccggt	gagtttttgt	acttttgaatt	gcccggaagag	catatcgagg	gcccgccgca	31980
cgcgctccgg	ctcacaccoc	aggttagagct	tacacgttagc	ctgattccggg	agtttaccaa	32040
gcgcgccctcg	ctagtgaggag	gggagcgggg	tccctgttgtt	ctgaccgtgg	ttgcacaatg	32100
tcttaacccct	ggattacatc	aagatcttat	tccattcaac	tacaataaaa	cacacaataa	32160
attacttact	taaaatcagt	cagcaaatct	ttgtccagct	tgatcagcat	caacctcttt	32220
ccctctccoc	aactctggta	tttcagcagc	cttttagctg	ggaactttct	ccaaagtcta	32280
aatgggatgt	caaatctctc	atgttcttgt	ccctccgcac	ccactatctt	catattgttg	32340
cagatgaaac	gcgccagacc	gtctgaagac	accttcaacc	ctgtgtaccc	atatgacacg	32400
gaaacgggcc	ctccaactgt	gcctttccct	acccctccct	ttgtgtogcc	aaatgggttc	32460
caagaagctc	cccccgaggt	gctttctttg	cgcttttcag	aacctttggt	tacctcacac	32520
ggcatgctgt	cgctaaaaat	gggcagcgcc	ctgtccctgg	atcaggcgag	caaccttaca	32580
tcaaatacaa	tactgttttc	tcaaccgcct	aaaaaaacaa	agtcacat	aaacttggaa	32640
acatccgcgc	cccttaccgt	cagctcaggc	gccctaacca	tggccacaac	ttcgctcttg	32700
gtggctctctg	acaacactct	taccatgcaa	tcacaagcac	cgctaaccgt	cgagaactca	32760
aaacttagca	ttgctaccaa	agagccactt	acagtggttg	atggaaaact	ggccctgcag	32820
acatcagccc	ccctctctgc	cactgataac	aacgcctcca	ctatcactgc	ctcaactcct	32880
cttactactg	caaatggtag	ctctgctgtt	acatggaaa	acccacttta	caacaacaa	32940
ggaaaacttg	ggctcaaaat	tggcggctct	ttgcaagtgg	ccacgcactc	acatgcaact	33000
acactaggtta	ctggttcagggt	gggttcagtt	catacaaat	tgtctacata	aaaagtata	33060
ggcgcaatag	ctgtttgatac	atctggcaac	atggaaacta	aaactggaga	tggcctctat	33120
gtggatagcg	ccggttcttaa	ccaaaacta	catattaatc	taataaccac	aaaaggcctt	33180
gcttttgaca	acacgcgaat	aaacattaac	gctggaaaag	ggttggaaat	tgaacacagc	33240
tctctcaacg	gaatcccat	aaaaacaaaa	atggatcag	gcatacaata	taaccacaa	33300
ggagctatgt	ttgcaaaaact	tggaaacaggc	ctcagttttg	acagctccgg	agccataaca	33360
atcgggcagca	taacaatagt	cagacttact	ctttggacaa	caccagaccc	atccccaaat	33420
tgcgaatgt	cttcagataa	agactgcgaag	ctaaactctg	cgctaacaaa	atgtggcgagt	33480
caaattttgg	gcactgtttc	agctttggca	gtatcaggta	atatggcctc	catcaatgga	33540
actctaagca	gtgtaaaact	ggttcttaga	tttgatgaca	acggagtgct	tatgtcaaat	33600
tcatcactgg	acaaaacagta	ttggaacttt	agaaacgggg	actccactaa	cggtcaacca	33660
tacacttatg	ctgttgggtt	tatgcocaaac	ctaaaagctt	accacaaaac	tcaaaagttaa	33720
actgcacaaa	gtaatatgtt	tagccagggtg	tatcttaatg	gtgacaaagt	taaacattgt	33780
cabtttacta	ttacgctaaa	tggaaacagat	gaaaccaacc	aagtaagcaa	atactcaata	33840
tcaattcagtt	ggctcctggaa	cagtgggacaa	tacactaatg	acaaatttgc	caccaattcc	33900
tatacctttct	ctcatactgc	ccaggaaataa	agaactcgta	acotgttgca	tgttatgttt	33960
caactgtttt	attttttcaat	tgcagaaaaat	ttcaagtcat	ttttcatcca	gtagtatagc	34020
ccccacacca	ctatagcttat	actaatcacc	gtaccttaat	caaaactoca	gaacctagtt	34080
attcaactcg	ccactccctc	cccaaccac	agagtacaca	gtcctttctc	cccggtctggc	34140
cttaaacagc	atcatatcat	gggttaacaga	catattctta	gtgtttatat	tccacacgtg	34200
ctctctgcga	gccaaacgct	catcagtgat	gttaataaac	tccccgggca	ctgcctgttaa	34260
gttcatgtgt	ctgttccagct	gctgagccac	aggtcgtctg	ccaacttgcg	gttgcctaac	34320
ggcgccggcaa	ggagaaagtcc	agcgcatacat	gggggttagag	tcaataatcg	gcatacagagt	34380
agggcggttg	gtctgcagca	gctcctcgaat	aaactctctgc	cgccgcgcct	cgctcctgca	34440
ggaaacacaa	atggcgagtgg	tctcctcagc	gttgtctgcg	accgcccgcg	gcataaggcg	34500
ctctgtctcc	cgggccagcag	agcgcaacct	gatctcactt	aagtcagcac	agtaactgca	34560
gcacagtaac	acaatatgtg	ttaaaaatccc	acagtgcaag	ggcgtgtatc	caaaagctcat	34620
ggcgggggacc	acaggaaccca	cgtaggcctc	ataccacaag	cgcaggtaga	ttaaagtggcg	34680
acccctcata	aaacgcgtgt	acataaacat	tacctctttt	ggcatgttgt	aattccaac	34740
ctcccggtac	catataaaaac	tctgatataa	catggcgcca	tcacacacca	tcttaaacca	34800
gctggcgcaaa	acotgcgcgc	cggtctatgca	ctgcaggggaa	ccgggactgg	aaacaatgaca	34860
gtggagagacc	caggactcgt	aacctatggat	catcatgtctc	gtcatgatata	caatgttggc	34920
acaaacacagg	cacacgtgca	tacacttctc	caggattaca	agctctctcc	gcgcagaac	34980
catatccagc	ggacaacccc	attcctgaat	cagcgttaaat	cccacactgc	agggagagacc	35040
tgcagagtaa	ctcaactgtgt	gcattgtcaa	agtggttacct	ctgggcagca	gggaagtgtc	35100
ctccagtatg	gtacgcggcg	tctctgtctc	aaaaggaggt	agacgatccc	tactgtacgg	35160

```

agtcgcgcga gacacccgag atcgtgttgg tegtagtgtc atgcacaatg gaacgcgcga 35220
cgtagtcata ttctcctaag caaaaccagg tgcgggcgtg acaaacagat ctgcgtctcc 35280
ggctcgcgcc cttagatcgc tctgtgtagt agttgttagt tatccactct ctcaaaagcat 35340
ccaggcgccc cctggcttcg ggttctatgt aaactccttc atgcgcgcgt cccttgataa 35400
catccaccac cgcagaataa gccacaccga gccaacctac acattcgttc tgcgagtcac 35460
acacggggag agcgggaaga gctgggaaga catgtttttt tttttatttc caaaagatta 35520
tccaaacctc caaatgaag atctattaa gaaacgcgtt cccctccggt ggctgtgtca 35580
aacctctacg ccaaaagaaca gataatggca tttgttagat gttgcacaat ggcttccaaa 35640
aggcaaacgg cccctcacgt caagtggagc taagggttaa acccttcagg gtgaatctcc 35700
tctataaaca tcccagcacc ttcaaccatg cccaaataat tctcatctcg ccacctcttc 35760
aatatacttc taagcaaatc ccgaatatta agtccggcca ttgtaaaaaa ctgcctccaga 35820
ggccctctca ccttcagcct caagcagcga atcatgattg caaaaattca ggttccctac 35880
agacctgtat aagattcaaa agcgggaacat taacaaaaat accgcgagtc cgtaggctcc 35940
ttcgcaggcg cagctgaaca taatcgtgca ggtctgcagc gaccagcgcg gcaactctcc 36000
cgccagggaac catgacaaaa gaacccacac tgattatgac acgcatactc ggaagctatgc 36060
taaccagcgt agcccgctag taagcttggt gcattggcgg cgtatataaa tgcgaaggtgc 36120
tgctcaaaaa atcaggcaaa gctcgcgcga aaaaagaaag cacatcgtag tcatgtctac 36180
gcagataaag gcaggtaagc tccgggaaca ccacagaaaa agacacacatt tttctctcaa 36240
acatgtctgc ggggtttctgc ataaacacaa aataaaataa caaaaaaaca ttttaaacatt 36300
agaagcgtgt cttacaacag gaaaaacac ccttataagc atagacgga ctaccggcat 36360
gcggcggtga cgtataaaaa actggtcacc gtgattaaaa agcaccaccc agacgtcttc 36420
gggtatgtcc ggagtcataa tgaagactc ggtaaacaca tcaggttgat tcacatcgtt 36480
cagtgctaaa aaggcagcga aatagcccg gggaatacat acccgcaggc tagagagcaa 36540
catcagccgc cccataggag gtatacaaaa attaatagga gaaaaaaca cataaacacc 36600
tgaaaaaccc tctcgctag gcaaaatagc accctccgc tcagaaacaa catacagcgc 36660
ttccacagcg gcagccataa cagtcagcct taccagtaaa aagaaaaacc tattaataaa 36720
acaccactcg acacggcacc agtccaatca gtccagtggt aaaaaaggcg caagtgcaga 36780
gcgagtatal ataggactaa aaatgacgt aacggttaaa gtccacaaa aacaccaga 36840
aaacgcgcag cgaacctacg cccagaacag aaagccaaaa aaccacaaac ttctcctaat 36900
cgtaactctc gttttccacc gttacgtcac ttccatttt aagaaacta caatctccaa 36960
cacatacaag ttactccgcc ctaaaaaccta cgtcacccgc ccgcttccca cgccccgcgc 37020
ccggtcacaa actccacccc ctcattatca tatggcttc aatccaaaat aaggtatat 37080
attgatgatg

```

<210> 5
 <211> 5955
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> NS cDNA sequence

<221> CDS
 <222> (1)...(5955)

```

<400> 5
atg gcg ccc atc acg gcc tac tcc caa cag acg cgg ggc cta ctt ggt      48
Met Ala Pro ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly
1          5          10          15

tgc atc atc act agc ctt aca ggc cgg gac aag aac cag gtc gac gga      96
Cys ile ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly
20          25          30

gag gtt cag gtg gtt tcc acc gca aca caa tcc ttc ctg gcg acc tgc      144

```

Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys	
35 40 45	
gtc aac ggc gtg tgt tgg acc gtt tac cat ggt gct ggc tca aag acc	192
Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr	
50 55 60	
tta gcc ggc cca aag ggg cca atc acc cag atg tac act aat gtg gac	240
Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp	
65 70 75 80	
cag gac ctc gtc ggc tgg cag gcg ccc ccc ggg gcg cgt tcc ttg aca	288
Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr	
85 90 95	
cca tgc acc tgt ggc agc tca gac ctt tac ttg gtc acg aga cat gct	336
Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala	
100 105 110	
gac gtc att ccg gtg cgc cgg cgg ggc gac agt agg ggg agc ctg ctc	384
Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu	
115 120 125	
tcc ccc agg cct gtc tcc tac ttg aag ggc tct tcg ggt ggt cca ctg	432
Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu	
130 135 140	
ctc tgc cct tcg ggg cac gct gtg ggc atc ttc cgg gct gcc gta tgc	480
Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys	
145 150 155 160	
acc cgg ggg gtt cgc aag gcg gtg gac ttt gtg ccc gta gag tcc atg	528
Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met	
165 170 175	
gaa act act atg cgg tct ccg gtc ttc acg gac aac tca tcc ccc ccg	576
Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro	
180 185 190	
gcc gta ccg cag tca ttt caa gtg gcc cac cta cac gct ccc act ggc	624
Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly	
195 200 205	
agc ggc aag agt act aaa gtg ccg gct gca tat gca gcc caa ggg tac	672
Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr	
210 215 220	
aag gtg ctc gtc ctc aat ccg tcc gtt gcc gct acc tta ggg ttt ggg	720
Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly	
225 230 235 240	
gcg tat atg tct aag gca cac ggt att gac ccc aac atc aga act ggg	768
Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly	
245 250 255	

gta agg acc att acc aca ggc gcc ccc gtc aca tac tct acc tat ggc Val Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly	816
260 265 270	
aag ttt ctt gcc gat ggt ggt tgc tct ggg ggc gct tat gac atc ata Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile	864
275 280 285	
ata tgt gat gag tgc cat tca act gac tgc act aca atc ttg ggc atc Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile	912
290 295 300	
ggc aca gtc ctg gac caa gcg gag acg gct gga gcg cgg ctt gtc gtg Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val	960
305 310 315 320	
ctc gcc acc gct acg cct ccg gga tgc gtc acc gtg cca cac cca aac Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn	1008
325 330 335	
atc gag gag gtg gcc ctg tct aat act gga gag atc ccc ttc tat ggc Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly	1056
340 345 350	
aaa gcc atc ccc att gaa gcc atc agg ggg gga agg cat ctc att ttc Lys Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe	1104
355 360 365	
tgt cat tcc aag aag aag tgc gac gag ctc gcc gca aag ctg tca ggc Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly	1152
370 375 380	
ctc gga atc aac gct gtg gcg tat tac ccg ggg ctc gat gtg tcc gtc Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val	1200
385 390 395 400	
ata cca act atc gga gac gtc gtt gtc gtg gca aca gac gct ctg atg Ile Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met	1248
405 410 415	
acg ggc tat acg ggc gac ttt gac tca gtg atc gac tgt aac aca tgt Thr Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys	1296
420 425 430	
gtc acc cag aca gtc gac ttc agc ttg gat ccc acc ttc acc att gag Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu	1344
435 440 445	
acg acg acc gtg cct caa gac gca gtg tgc cgc tgc cag cgg cgg ggt Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Gly	1392
450 455 460	
agg act ggc agg ggt agg aga ggc atc tac agg ttt gtg act ccg gga Arg Thr Gly Arg Gly Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly	1440
465 470 475 480	

gaa cgg ccc tcg gcc atg ttc gat tcc tcg gtc ctg tgt gag tgc tat Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr 485 490 495	1488
gac gcg gcc tgt gct tgg tac gag ctc acc ccc gcc gag acc tcg gtt Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val 500 505 510	1536
agg ttg cgg gcc tac ctg aac aca cca ggg ttg ccc gtt tgc cag gac Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp 515 520 525	1584
cac ctg gag ttc tgg gag agt gtc ttc aca gcc ctc acc cac ata gat His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp 530 535 540	1632
gca cac ttc ttg tcc cag acc aag cag gca gga gac aac ttc ccc tac Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr 545 550 555 560	1680
ctg gta gca tac caa gcc acg gtg tgc gcc agg gct cag gcc cca cct Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro 565 570 575	1728
cca tca tgg gat caa atg tgg aag tgt ctc ata cgg ctg aaa cct acg Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr 580 585 590	1776
ctg cac ggg cca aca ccc ttg ctg tac agg ctg gga gcc gtc caa aat Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn 595 600 605	1824
gag gtc acc ctc acc cac ccc ata acc aaa tac atc atg gca tgc atg Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met 610 615 620	1872
tcg gct gac ctg gag gtc gtc act agc acc tgg gtg ctg gtg gcc gga Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly 625 630 635 640	1920
gtc ctt gca gct ctg gcc gcg tat tgc ctg aca aca gcc agt gtg gtc Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Thr Thr Gly Ser Val Val 645 650 655	1968
att gtg ggt agg att atc ttg tcc ggg agg ccg gct att gtt ccc gac Ile Val Gly Arg Ile Ile Leu Ser Gly Arg Pro Ala Ile Val Pro Asp 660 665 670	2016
agg gag ttt ctc tac cag gag ttc gat gaa atg gaa gag tgc gcc tcg Arg Glu Phe Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ala Ser 675 680 685	2064
cac ctc cct tac atc gag cag gga atg cag ctc gcc gag caa ttc aag	2112

His Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe Lys 690 695 700	
cag aaa gcg ctc ggg tta ctg caa aca gcc acc aaa caa gcg gag gct Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu Ala 705 710 715 720	2160
gct gct ccc gcg gtg gag tcc aag tgg cga gcc ctt gag aca ttc tgg Ala Ala Pro Val Val Glu Ser Lys Trp Arg Ala Leu Glu Thr Phe Trp 725 730 735	2208
gcg aag cac atg tgg aat ttc atc agc ggg ata cag tac tta gca gcc Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly 740 745 750	2256
tta tcc act ctg cct ggg aac ccc gca ata gca tca ttg atg gca ttc Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe 755 760 765	2304
aca gcc tct atc acc agc ccg ctc acc acc caa agt acc ctc ctg ttt Thr Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln Ser Thr Leu Leu Phe 770 775 780	2352
aac atc ttg ggg ggg tgg gtg gct gcc caa ctc gcc ccc ccc agc gcc Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser Ala 785 790 795 800	2400
gct tcg gct ttc gtg ggc gcc ggc atc gcc ggt gcg gct gtt ggc agc Ala Ser Ala Phe Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly Ser 805 810 815	2448
ata ggc ctt ggg aag gtg ctt gtg gac att ctg gcg ggt tat gga gca Ile Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala 820 825 830	2496
gga gtg gcc ggc gcg ctc gtg gcc ttc aag gtc atg agc ggc gag atg Gly Val Ala Gly Ala Leu Val Ala Phe Lys Val Met Ser Gly Glu Met 835 840 845	2544
ccc tcc acc gag gac ctg gtc aat cta ctt cct gcc atc ctc tct cct Pro Ser Thr Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro 850 855 860	2592
ggc gcc ctg gtc gtc ggg gtc gtg tgt gca gca ata ctg cgt cga cac Gly Ala Leu Val Val Gly Val Val Cys Ala Ala Ile Leu Arg Arg His 865 870 875 880	2640
gtg ggt ccg gga gag ggg gct gtg cag tgg atg aac cgg ctg ata gcg Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala 885 890 895	2688
ttc gcc tcg cgg ggt aat cat gtt tcc ccc acg cac tat gtg cct gag Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu 900 905 910	2736

agc gac gcc gca gcg cgt gtt act cag atc ctc tcc agc ctt acc atc Ser Asp Ala Ala Ala Arg Val Thr Gln Ile Leu Ser Ser Leu Thr Ile 915 920 925	2784
act cag ctg ctg aaa agg ctc cac cag tgg att aat gaa gac tgc tcc Thr Gln Leu Leu Lys Arg Leu His Gln Trp Ile Asn Glu Asp Cys Ser 930 935 940	2832
aca ccg tgt tcc gcc tgc tgg cta agg gat gtt tgg gac tgg ata tgc Thr Pro Cys Ser Gly Ser Trp Leu Arg Asp Val Trp Asp Trp Ile Cys 945 950 955 960	2880
acg gtg ttg act gac ttc aag acc tgg ctc cag tcc aag ctc ctg ccg Thr Val Leu Thr Asp Phe Lys Thr Trp Leu Gln Ser Lys Leu Leu Pro 965 970 975	2928
cag cta ccg gga gtc cct ttt ttc tgc tgc caa cgc ggg tac aag gga Gln Leu Pro Gly Val Pro Phe Phe Ser Cys Gln Arg Gly Tyr Lys Gly 980 985 990	2976
gtc tgg ccg gga gac gcc atc atg caa acc acc tgc cca tgt gga gca Val Trp Arg Gly Asp Gly Ile Met Gln Thr Thr Cys Pro Cys Gly Ala 995 1000 1005	3024
cag atc acc gga cat gtc aaa aac ggt tcc atg agg atc gtc ggg cct Gln Ile Thr Gly His Val Lys Asn Gly Ser Met Arg Ile Val Gly Pro 1010 1015 1020	3072
aag acc tgc agc aac acg tgg cat gga aca ttc ccc atc aac gca tac Lys Thr Cys Ser Asn Thr Trp His Gly Thr Phe Pro Ile Asn Ala Tyr 1025 1030 1035 1040	3120
acc acg gcc ccc tgc aca ccc tct cca gcg cca aac tat tct agg gcg Thr Thr Gly Pro Cys Thr Pro Ser Pro Ala Pro Asn Tyr Ser Arg Ala 1045 1050 1055	3168
ctg tgg ccg gtg gcc gct gag gag tac gtg gag gtc acg ccg gtg ggg Leu Trp Arg Val Ala Ala Glu Glu Tyr Val Glu Val Thr Arg Val Gly 1060 1065 1070	3216
gat ttc cac tac gtg acg gcc atg acc act gac aac gta aag tgc cca Asp Phe His Tyr Val Thr Gly Met Thr Thr Asp Asn Val Lys Cys Pro 1075 1080 1085	3264
tgc cag gtt ccg gct cct gaa ttc ttc acg gag gtg gac gga gtg ccg Cys Gln Val Pro Ala Pro Glu Phe Phe Thr Glu Val Asp Gly Val Arg 1090 1095 1100	3312
ttg cac agg tac gct ccg gcg tgc agg cct ctc cta ccg gag gag gtt Leu His Arg Tyr Ala Pro Ala Cys Arg Pro Leu Leu Arg Glu Glu Val 1105 1110 1115 1120	3360
aca ttc cag gtc ggg ctc aac caa tac ctg gtt ggg tca cag cta cca Thr Phe Gln Val Gly Leu Asn Gln Tyr Leu Val Gly Ser Gln Leu Pro 1125 1130 1135	3408

tgc gag ccc gaa ccg gat gta gca gtg ctc act tcc atg ctc acc gac Cys Glu Pro Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp 1140 1145 1150	3456
ccc tcc cac atc aca gca gaa acg gct aag cgt agg ttg gcc agg ggg Pro Ser His Ile Thr Ala Glu Thr Ala Lys Arg Arg Leu Ala Arg Gly 1155 1160 1165	3504
tct ccc ccc tcc ttg gcc agc tct tca gct agc cag ttg tct gcg cct Ser Pro Pro Ser Leu Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro 1170 1175 1180	3552
tcc ttg aag gcg aca tgc act acc cac cat gtc tct ccg gac gct gac Ser Leu Lys Ala Thr Cys Thr Thr His His Val Ser Pro Asp Ala Asp 1185 1190 1195 1200	3600
ctc atc gag gcc aac ctc ctg tgg cgg cag gag atg ggc ggg aac atc Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile 1205 1210 1215	3648
acc cgc gtg gag tgc gag aac aag gtg gta gtc ctg gac tct ttc gac Thr Arg Val Glu Ser Glu Asn Lys Val Val Val Leu Asp Ser Phe Asp 1220 1225 1230	3696
ccg ctt cga gcg gag gag gat gag agg gaa gta tcc gtt ccg gcg gag Pro Leu Arg Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala Glu 1235 1240 1245	3744
atc ctg cgg aaa tcc aag aag ttc ccc gca gcg atg ccc atc tgg gcg Ile Leu Arg Lys Ser Lys Lys Phe Pro Ala Ala Met Pro Ile Trp Ala 1250 1255 1260	3792
cgc ccg gat tac aac cct cca ctg tta gag tcc tgg aag gac ccg gac Arg Pro Asp Tyr Asn Pro Pro Leu Leu Glu Ser Trp Lys Asp Pro Asp 1265 1270 1275 1280	3840
tac gtc cct ccg gtg gtg cac ggg tgc ccg ttg cca cct atc aag gcc Tyr Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Ile Lys Ala 1285 1290 1295	3888
cct cca ata cca cct cca cgg aga aag agg acg gtt gtc cta aca gag Pro Pro Ile Pro Pro Pro Arg Arg Lys Arg Thr Val Val Leu Thr Glu 1300 1305 1310	3936
tcc tcc gtg tct tct gcc tta gcg gag ctc gct act aag acc ttc ggc Ser Ser Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe Gly 1315 1320 1325	3984
agc tcc gaa tca tgc gcc gtc gac agc ggc acg gcg acc gcc ctt cct Ser Ser Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Leu Pro 1330 1335 1340	4032
gac cag gcc tcc gac gac ggt gac aaa gga tcc gac gtt gag tgc tac	4080

Asp Gln Ala Ser Asp Asp Gly Asp Lys Gly Ser Asp Val Glu Ser Tyr 1345 1350 1355 1360	
tcc tcc atg ccc ccc ctt gag ggg gaa cgg ggg gac ccc gat ctc agt Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser 1365 1370 1375	4128
gac ggg tct tgg tct acc gtg agc gag gaa gct agt gag gat gtc gtc Asp Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val Val 1380 1385 1390	4176
tgc tgc tca atg tcc tac aca tgg aca ggc gcc ttg atc acg cca tgc Cys Cys Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys 1395 1400 1405	4224
gct gcg gag gaa agc aag ctg ccc atc aac gcg ttg agc aac tct ttg Ala Ala Glu Glu Ser Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu 1410 1415 1420	4272
ctg cgc cac cat aac atg gtt tat gcc aca aca tct cgc agc gca ggc Leu Arg His His Asn Met Val Tyr Ala Thr Thr Ser Arg Ser Ala Gly 1425 1430 1435 1440	4320
ctg cgg cag aag aag gtc acc ttt gac aga ctg caa gtc ctg gac gac Leu Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Asp 1445 1450 1455	4368
cac tac cgg gac gtg ctc aag gag atg aag gcg aag gcg tcc aca gtt His Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val 1460 1465 1470	4416
aag gct aaa ctc cta tcc gta gag gaa gcc tgc aag ctg acg ccc cca Lys Ala Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Lys Thr Pro Pro 1475 1480 1485	4464
cat tcg gcc aaa tcc aag ttt ggc tat ggg gca aag gac gtc cgg aac His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn 1490 1495 1500	4512
cta tcc agc aag gcc gtt aac cac atc cac tcc gtg tgg aag gac ttg Leu Ser Ser Lys Ala Val Asn His Ile His Ser Val Trp Lys Asp Leu 1505 1510 1515 1520	4560
ctg gaa gac act gtg aca cca att gac acc acc atc atg gca aaa aat Leu Glu Asp Thr Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn 1525 1530 1535	4608
gag gtt ttc tgt gtc caa cca gag aaa gga ggc cgt aag cca gcc cgc Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg 1540 1545 1550	4656
ctt atc gta ttc cca gat ctg gga gtc cgt gta tgc gag aag atg gcc Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala 1555 1560 1565	4704

ctc tat gat gtg gtc tcc acc ctt cct cag gtc gtg atg ggc tcc tca Leu Tyr Asp Val Val Ser Thr Leu Pro Gln Val Val Met Gly Ser Ser 1570 1575 1580	4752
tac gga ttc cag tac tct cct ggg cag cga gtc gag ttc ctg gtg aat Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn 1585 1590 1595 1600	4800
acc tgg aaa tca aag aaa aac ccc atg ggc ttt tca tat gac act cgc Thr Trp Lys Ser Lys Lys Asn Pro Met Gly Phe Ser Tyr Asp Thr Arg 1605 1610 1615	4848
tgt ttc gac tca acg gtc acc gag aac gac atc cgt gtt gag gag tca Cys Phe Asp Ser Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu Ser 1620 1625 1630	4896
att tac caa tgt tgt gac ttg gcc ccc gaa gcc aga cag gcc ata aaa Ile Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Lys 1635 1640 1645	4944
tcg ctc aca gag cgg ctt tat atc ggg ggt cct ctg act aat tca aaa Ser Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys 1650 1655 1660	4992
ggg cag aac tgc ggt tat cgc cgg tgc cgc gcg agc ggc gtg ctg acg Gly Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr 1665 1670 1675 1680	5040
act agc tgc ggt aac acc ctc aca tgt tac ttg aag gcc tct gca gcc Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala 1685 1690 1695	5088
tgt cga gct gcg aag ctc cag gac tgc acg atg ctc gtg aac gga gac Cys Arg Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Asn Gly Asp 1700 1705 1710	5136
gac ctt gtc gtt atc tgt gaa agc gcg gga acc caa gag gac gcg gcg Asp Leu Val Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala 1715 1720 1725	5184
agc cta cga gtc ttc acg gag gct atg act agg tac tct gcc ccc ccc Ser Leu Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro 1730 1735 1740	5232
ggg gac ccg ccc caa cca gaa tac gac ttg gag ctg ata aca tca tgt Gly Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys 1745 1750 1755 1760	5280
tcc tcc aat gtg tgc gtc gcc cac gat gca tca ggc aaa agg gtg tac Ser Ser Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr 1765 1770 1775	5328
tac ctc acc cgt gat ccc acc acc ccc ctc gca cgg gct gcg tgg gaa Tyr Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu 1780 1785 1790	5376

aca gct aga cac act cca gtt aac tcc tgg cta ggc aac att atc atg 5424
 Thr Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met
 1795 1800 1805

tat gcg ccc act ttg tgg gca agg atg att ctg atg act cac ttc ttc 5472
 Tyr Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe
 1810 1815 1820

tcc atc ctt cta gca cag gag caa ctt gaa aaa gcc ctg gac tgc cag 5520
 Ser Ile Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln
 1825 1830 1835 1840

atc tac ggg gcc tgt tac tcc att gag cca ctt gac cta cct cag atc 5568
 Ile Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile
 1845 1850 1855

att gaa cga ctc cat ggc ctt agc gca ttt tca ctc cat agt tac tct 5616
 Ile Glu Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser
 1860 1865 1870

cca ggt gag atc aat agg gtg gct tca tgc ctc agg aaa ctt ggg gta 5664
 Pro Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val
 1875 1880 1885

cca ccc ttg cga gtc tgg aga cat cgg gcc agg agc gtc cgc gct agg 5712
 Pro Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg
 1890 1895 1900

cta ctg tcc cag ggg ggg agg gcc gcc act tgt ggc aag tac ctc ttc 5760
 Leu Leu Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe
 1905 1910 1915 1920

aac tgg gca gtg aag acc aaa ctc aaa ctc act cca atc ccg gct gcg 5808
 Asn Trp Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala
 1925 1930 1935

tcc cag ctg gac ttg tcc ggc tgg ttc gtt gct ggt tac agc ggg gga 5856
 Ser Gln Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly
 1940 1945 1950

gac ata tat cac agc ctg tct cgt gcc cga ccc cgc tgg ttc atg ctg 5904
 Asp Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu
 1955 1960 1965

tgc cta ctc cta ctt tct gta ggg gta ggc atc tac ctg ctc ccc aac 5952
 Cys Leu Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn
 1970 1975 1980

cga 5955
 Arg
 1985

<210> 6
 <211> 1984

<212> PRT

<213> Artificial Sequence

<220>

<223> NS sequence

<400> 6

```

Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys
 1           5           10           15
Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu
 20           25           30
Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys Val
 35           40           45
Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr Leu
 50           55           60
Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp Gln
 65           70           75
Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr Pro
 85           90           95
Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp
100           105           110
Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser
115           120           125
Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu
130           135           140
Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys Thr
145           150           155
Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met Glu
165           170           175
Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala
180           185           190
Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser
195           200           205
Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys
210           215           220
Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala
225           230           235
Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val
245           250           255
Arg Thr Ile Thr Thr Gly Ala Pro Val Thr Tyr Ser Thr Tyr Gly Lys
260           265           270
Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile
275           280           285
Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile Gly
290           295           300
Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu
305           310           315
Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile
325           330           335
Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly Lys
340           345           350
Ala Ile Pro Ile Glu Ala Ile Arg Gly Gly Arg His Leu Ile Phe Cys
355           360           365
His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly Leu
370           375           380

```

Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile
 385 390 395 400
 Pro Thr Ile Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met Thr
 405 410 415
 Gly Tyr Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val
 420 425 430
 Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr
 435 440 445
 Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly Arg
 450 455 460
 Thr Gly Arg Gly Arg Arg Gly Ile Tyr Arg Phe Val Thr Pro Gly Glu
 465 470 475 480
 Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp
 485 490 495
 Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val Arg
 500 505 510
 Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His
 515 520 525
 Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp Ala
 530 535 540
 His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu
 545 550 555 560
 Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro
 565 570 575
 Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu
 580 585 590
 His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu
 595 600 605
 Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met Ser
 610 615 620
 Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val
 625 630 635 640
 Leu Ala Ala Leu Ala Tyr Cys Leu Thr Thr Gly Ser Val Val Ile
 645 650 655
 Val Gly Arg Ile Ile Leu Ser Gly Arg Pro Ala Ile Val Pro Asp Arg
 660 665 670
 Glu Phe Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ala Ser His
 675 680 685
 Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe Lys Gln
 690 695 700
 Lys Ala Leu Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu Ala Ala
 705 710 715 720
 Ala Pro Val Val Glu Ser Lys Trp Arg Ala Leu Glu Thr Phe Trp Ala
 725 730 735
 Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu
 740 745 750
 Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr
 755 760 765
 Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln Ser Thr Leu Leu Phe Asn
 770 775 780
 Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro Ser Ala Ala
 785 790 795 800
 Ser Ala Phe Val Gly Ala Gly Ile Ala Gly Ala Ala Val Gly Ser Ile
 805 810 815
 Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala Gly

	820		825		830
Val	Ala	Gly	Ala	Leu	Val
	835		840		845
Ser	Thr	Glu	Asp	Leu	Val
	850		855		860
Ala	Leu	Val	Val	Gly	Val
	865		870		875
Gly	Pro	Gly	Glu	Gly	Ala
	885		890		895
Ala	Ser	Arg	Gly	Asn	His
	900		905		910
Asp	Ala	Ala	Arg	Val	Thr
	915		920		925
Gln	Leu	Leu	Lys	Arg	Leu
	930		935		940
Pro	Cys	Ser	Gly	Ser	Trp
	945		950		955
Val	Leu	Thr	Asp	Phe	Lys
	965		970		975
Leu	Pro	Gly	Val	Pro	Phe
	980		985		990
Trp	Arg	Gly	Asp	Gly	Ile
	995		1000		1005
Ile	Thr	Gly	His	Val	Lys
	1010		1015		1020
Thr	Cys	Ser	Asn	Thr	Trp
	1025		1030		1035
Thr	Gly	Pro	Cys	Thr	Pro
	1045		1050		1055
Trp	Arg	Val	Ala	Ala	Glu
	1060		1065		1070
Phe	His	Tyr	Val	Thr	Gly
	1075		1080		1085
Gln	Val	Pro	Ala	Pro	Glu
	1090		1095		1100
His	Arg	Tyr	Ala	Pro	Ala
	1105		1110		1115
Phe	Gln	Val	Gly	Leu	Asn
	1125		1130		1135
Glu	Pro	Glu	Pro	Asp	Val
	1140		1145		1150
Ser	His	Ile	Thr	Ala	Glu
	1155		1160		1165
Pro	Pro	Ser	Leu	Ala	Ser
	1170		1175		1180
Leu	Lys	Ala	Thr	Cys	Thr
	1185		1190		1195
Ile	Glu	Ala	Asn	Leu	Leu
	1205		1210		1215
Arg	Val	Glu	Ser	Glu	Asn
	1220		1225		1230
Leu	Arg	Ala	Glu	Glu	Asp
	1235		1240		1245
Leu	Arg	Lys	Ser	Lys	Lys
	1250		1255		1260

Pro Asp Tyr Asn Pro Pro Leu Leu Glu Ser Trp Lys Asp Pro Asp Tyr
 1265 1270 1275 1280
 Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Ile Lys Ala Pro
 1285 1290 1295
 Pro Ile Pro Pro Pro Arg Arg Lys Arg Thr Val Val Leu Thr Glu Ser
 1300 1305 1310
 Ser Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe Gly Ser
 1315 1320 1325
 Ser Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Leu Pro Asp
 1330 1335 1340
 Gln Ala Ser Asp Asp Gly Asp Lys Gly Ser Asp Val Glu Ser Tyr Ser
 1345 1350 1355 1360
 Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp
 1365 1370 1375
 Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val Val Cys
 1380 1385 1390
 Cys Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys Ala
 1395 1400 1405
 Ala Glu Glu Ser Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu
 1410 1415 1420
 Arg His His Asn Met Val Tyr Ala Thr Thr Ser Arg Ser Ala Gly Leu
 1425 1430 1435 1440
 Arg Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Asp His
 1445 1450 1455
 Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val Lys
 1460 1465 1470
 Ala Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro His
 1475 1480 1485
 Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn Leu
 1490 1495 1500
 Ser Ser Lys Ala Val Asn His Ile His Ser Val Trp Lys Asp Leu Leu
 1505 1510 1515 1520
 Glu Asp Thr Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu
 1525 1530 1535
 Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu
 1540 1545 1550
 Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu
 1555 1560 1565
 Tyr Asp Val Val Ser Thr Leu Pro Gln Val Val Met Gly Ser Ser Tyr
 1570 1575 1580
 Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn Thr
 1585 1590 1595 1600
 Trp Lys Ser Lys Lys Asn Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys
 1605 1610 1615
 Phe Asp Ser Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu Ser Ile
 1620 1625 1630
 Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Lys Ser
 1635 1640 1645
 Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys Gly
 1650 1655 1660
 Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr
 1665 1670 1675 1680
 Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ser Ala Ala Cys
 1685 1690 1695

Arg Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Asn Gly Asp Asp
 1700 1705 1710
 Leu Val Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala Ser
 1715 1720 1725
 Leu Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly
 1730 1735 1740
 Asp Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser
 1745 1750 1755 1760
 Ser Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr Tyr
 1765 1770 1775
 Leu Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr
 1780 1785 1790
 Ala Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Tyr
 1795 1800 1805
 Ala Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser
 1810 1815 1820
 Ile Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln Ile
 1825 1830 1835 1840
 Tyr Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile Ile
 1845 1850 1855
 Glu Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro
 1860 1865 1870
 Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val Pro
 1875 1880 1885
 Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu
 1890 1895 1900
 Leu Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe Asn
 1905 1910 1915 1920
 Trp Ala Val Lys Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala Ser
 1925 1930 1935
 Gln Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Ser Gly Gly Asp
 1940 1945 1950
 Ile Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu Cys
 1955 1960 1965
 Leu Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
 1970 1975 1980

<210> 7

<211> 4909

<212> DNA

<213> Artificial Sequence

<220>

<223> pVLJ nucleic acid

<400> 7

tcgcgcgttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacggta	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcaggggcgc	tcagcgggtg	120
ttggcgggtg	tcggggctcg	cttaactatg	cggcacacga	gcagattgta	ctggagatgc	180
accatatgcy	ggtgaaata	cgcacacagat	gcgtaaggag	aaaataccgc	atcagattgg	240
ctattggcca	ttgcatacgt	tgatccata	tcataatag	tacattata	ttggctcatg	300
tcacaacatta	cgcacatggt	gacattgatt	attgactagt	tattaatagt	aatacaattac	360
ggggctcaata	gttcatagcc	catatatgga	gttccgcgtt	acataactta	cggtaaatgg	420
ccgcctggcg	tgaccgccca	acgacccccc	cccattgacg	tcaataatga	cgtatgttcc	480
catagtaacg	ccaataggga	ctttccattg	acgtcaatgg	gtggagattt	tacggtaaac	540

tgcccaactg	gcagtagatc	aagtgtatca	tatgccaaat	agccccccta	ttgacgtcaa	600
tgacggtaaa	tgcccgccgc	ggcattatgc	ccagtagatg	accttatggg	accttccatc	660
ttggcagtag	atctacgtat	tagtcatcgc	tattaccatg	tgatgcgggt	tttggcagta	720
catcaatggg	cgttgtagac	gggtttagtc	acgggggatt	ccaagtctcc	accccatbga	780
cgtcaatggg	agtttgtttt	ggcaccacaa	tcaacgggac	tttccaaat	gtcgtataca	840
ctccgcccac	ttgacgcmaa	tggtgggttag	gcgtgtacgg	tggtgggtct	atataagcag	900
agctcgttta	gtgaacgcgc	agatcgccgt	gagacgcocg	ccacgctgtt	ttgacctcca	960
tagaagacac	cggaaccgat	ccagcctccg	cgccgggaaa	cggtgacatt	gaacgcggat	1020
tcccgcgtcc	aagatgtacg	taagtacocg	ctatagactc	tataggccaca	cccttctggc	1080
tcttatgcac	gctatactgt	ttttggcgtt	gggocctata	acccccgctt	cccttatgcta	1140
taggttgatgg	tatagcttag	ccctataggtg	tggtgtattg	accattattg	accactcccc	1200
tattgttgac	gatactttcc	attactaatc	cataacatgg	ctcttttgca	caactatctc	1260
tattgtctat	atgccaatac	ctctgtccctc	agagactgac	acggactctg	tatttttaca	1320
ggatgggggtc	ccatttatta	tttacaacat	cacatatata	acaacgcggt	cccccggtcc	1380
cgcatgtttt	attaaacata	cggtggggtc	tccacgcgaa	tctcgggtac	gtgttccggc	1440
catgggctct	tctccggtag	cgccggagct	tccacatccg	agcccttggtc	ccatgcctcc	1500
agcggctcat	ggctgcctcg	cagctccctg	ctcctaacag	tggaggccag	acttaggcac	1560
agcacaatgc	ccaccaccac	cagtgtgcgc	cacaaggccg	tggcggtagg	gtatgtgtct	1620
gaataaagcg	gtggagattg	ggctgcgcag	gctgacgcag	atgggaagct	taagggtcag	1680
gcagagaaga	atggaagcag	ctgagtttgt	gtattctgat	aagagtccaga	ggtaactccc	1740
gttgcggctg	tgtttaacgt	ggagggcagt	gtagtctgag	cagtagctcg	tgcctgcggc	1800
cgccgcacca	gcataaatag	ctgacagact	aacagactgt	tcccttccat	gggtcttttc	1860
tcagctccac	gtcccttagat	ctaggtagca	gatatacaga	ttcagtcgac	agggcgcggg	1920
atctgtctgt	ctctctagtt	gccagccatc	tggtgtttgc	ccctccccgc	tgcccttccct	1980
gacccctggaa	ggtgcaccct	ccactgtccct	ttcctaataa	aatgaggaaa	tgtcatcgca	2040
ttgtctgagt	aggtgtccatt	ctattctggg	gggtgggggtg	gggcaggaca	gcaaggggga	2100
ggatggggaa	gacaatagca	ggcatgctgg	ggatgcgggtg	ggctctatgg	ccgtctccgc	2160
caggtgtctga	agaattgacc	cggttccctc	tgggccagaa	agaagcaggc	acatccccct	2220
ctctgtgaca	caacctgtcc	acgccccctg	ttcttagttc	cagccccact	cataggacac	2280
tcatagtagc	ggagggctcc	gcccttaact	ccaccgccta	aagtacttgg	agcggctctc	2340
ccctccctca	tcagcccccac	aaacccaaac	tagctcccaa	gagtgggaag	aaattaaagc	2400
aagataggtct	attaaagtca	gagggagaga	aaatgcctcc	aacatgtgag	gaagttaatga	2460
gagaaattcat	agaatttctt	ccgcttctct	gctcacttgc	gcgtctgcct	cggctgtctg	2520
gctgcggcga	gggttatcgac	ctcaactcaa	ggcgttaata	cggttatcca	cagtaacagg	2580
ggataaegca	ggaaaagaaca	tgtgagcaaa	aggccagaca	aaggccagga	accgtataaaa	2640
ggccgcgttg	ctggcgtttt	tccataggtc	cggccccctc	gagcagcacc	acaaaaatcg	2700
acgctcaagt	cagaggttggc	gaaccccgac	aggactataa	agataaccag	cggtttccccc	2760
tggaaagctcc	ctcgtgcgct	ctcctgttcc	gaccttgccg	cttaccggat	acctgtccgc	2820
ctttctccct	tggggaaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	2880
gggtgtaggtc	gttgcgtcca	agctgggctg	tgtgcacgaa	ccccccgttc	agcccgacgc	2940
ctgcgcctta	tccggttaact	atcgttctga	gtccaaaccg	gtaagacacg	acctatcgcc	3000
actgggcagca	gccactggta	acaggaattag	cagagcgagc	tatgtaggcg	gtgtcataga	3060
gtctcttgtag	tgggtgcctca	actacggcta	cactagaaga	acagatattg	gtatcttgccg	3120
tctgtcgaag	ccagttacct	tccgaaaaag	agtttgtagc	ttctgtatcc	gcacaaacac	3180
ccacgcgggt	agcgggtggtt	ttttgtgttg	caagcagcag	attacgcgca	gaaaaaaagg	3240
atctcaagaa	gacctcttga	tctttctcac	ggggtctgac	gctcagttga	acgaaatactc	3300
acgttaaggg	atttttggtca	tgagattatc	aaaaaggatc	tccacttaga	tctttttaaa	3360
ttaaaaatga	agtttttaaat	caactctaaag	tatatatgag	taaaacttgg	ctgacagtta	3420
ccaatgctta	atcagtgagg	caoctatctc	agcgatctgt	ctatttctgt	catccatagt	3480
tgccctgactc	ggggggggggg	ggcgctgagg	tctgcctcgt	gaagaagggtg	gtctgacttc	3540
ataccaggcc	tgaatgcgcc	catcatccag	ccagaaagtg	agggagccac	gggtgtatgag	3600
agctttgtgtg	taagtgagacc	agtttggtgat	tttgaaactt	tgcttttgcca	cggaaacggctc	3660
tcgctgtctgc	ggagagatgcg	tgatctgatc	cttcaactca	gcnaaaagttc	gatttatcca	3720
acaaagccgc	cgctccctca	agtcagcgta	atgctctgcc	agttgtacaa	ccaatataacc	3780
aattctgatt	agaaaaactc	atcgagcatc	aaatgaact	gcaatttatt	catatcagga	3840

ttatcaaatc	catatgtttg	aaaaagccgt	ttctgtaatg	aaggagaaaa	ctcaccgagg	3900
cagttccata	ggatggcaag	atcctgggat	cggtctgcga	ctccgactcg	tcacaactca	3960
atcacaacct	taaatttccc	ctcgtc3aaa	ataaggttat	caagtggaga	atcaccatga	4020
gtgacgactg	aatccgggtg	gaatggc3aa	agcttatgca	ttcttttcca	gacttggtca	4080
acaggccagc	catcacgctc	gtcatcaaaa	tcactcgcat	caacccaacc	gttattcatt	4140
ctgtgattgg	cctgaacgag	acgaataacg	cgatcgtgtg	taaaaggaga	attacaacaa	4200
ggaaatcgaa	gcaaacggcg	caggaaacat	gccagcgcat	caacaataat	ttcaccgtga	4260
tcaggatatt	cttctaatac	ctggaaatgt	gttttccggc	ggatcgcgat	ggcgagttaac	4320
catgcatcat	caggagtagc	gataaaatgc	ttgatggctg	gaagaggcat	aaattccgtc	4380
agccagttta	gtctgaccat	ctcatctgtg	acatcatatg	caacgtatcc	tttgccatgt	4440
ttcagaaaca	actctggcgc	atcgggcttc	ccatataaat	gatagattgt	cgcacattgt	4500
tgcccgacat	tcagcgagcg	ccattttatac	ccatataaat	cagcatccat	gttggaaatt	4560
aatcgcgccc	tcagacaaga	cgtttcccggt	tgaatatggc	tcataacacc	cttggtatta	4620
ctgtttatgt	aagcagacag	ttttatttgtt	catgatgata	tattttttatc	ttgtgcaatg	4680
taacatcaga	gatttttgaga	cacaaacgtgg	ctttccccc	ccccccatta	tttgaagcatt	4740
tatcagggtt	attgtctcat	gagcggatac	atattttgaat	gtattttagaa	aaatacaaaa	4800
atagggggtc	cgcgcacatt	tcgccgaaaa	gtgccacctg	acgtctaaga	aaccatttatt	4860
atcatgacat	taacctataa	aaataggcgt	atcacgagcg	ccatttcgtc		4909

<210> 8

<211> 35935

<212> DNA

<213> Adenovirus serotype 6

<400> 8

catcatcaat	aatataacct	atatttgatt	gaagccaata	tgataatgag	ggggtggagt	60
ttgtgacgtg	gcgcggggcg	tgggaaacgg	gcgggtgacg	tagtagtgtg	gcggaagtgt	120
gatgttgcaa	gtgtgcggga	acacatgtaa	gcgacggatg	tgccaaaagt	gacgtttttg	180
gtgtgcggcg	gtgtacacag	gaagtacaaa	ttttcgcgcg	gttttaggcg	gatgttgtag	240
taaaatttggg	cgtaacccag	taagatttgg	ccattttcgc	gggaaaactg	aataagagga	300
agtgaattct	gaataatttt	gtgttaactca	tagcgcgtaa	tatttgtcta	gggcgcgggg	360
gactttgacc	gtttactgtg	agactcgccc	aggtgttttt	ctcaggtgtt	ttccgcgttc	420
cgggtcaaa	ttggcggttt	attattatag	tcagctgacg	tgtagtgtat	ttataccggg	480
tgagttctct	aaaggccacc	tccttgagtg	cagcagagtg	agttttctct	tcgcagccgc	540
tcgcacacgc	ggactgaaaa	tgagacatat	tatctgccac	ggaggtgtta	taccgcaaga	600
aaatggccgc	agttcttttg	accagctgat	cgaagaggtg	ctggctgata	attcttccacc	660
tcctagccat	ttgaaaccac	ctacccttca	cgaactgtat	gatttagacg	tgacgcggcc	720
cgaagatccc	aacgagggag	cgggttcgca	gatttttccc	gactctgtaa	tgttggcggt	780
gcaggaaggg	attgacttac	tcacttttcc	gcgcgcggcc	ggtctccggg	agccgcctca	840
cttttccggg	cagcccgacg	agccggagca	gagagccttg	ggtccgggtt	ctatgccaac	900
ccctgtaccg	gaggtgatcg	atcttaacct	ccacgaggct	ggctttccac	ccagtgacga	960
caggagtgaa	gagggtaggg	agtttgtgtt	agattatgtg	gagcaccccg	ggcacggttg	1020
caggtctctg	catatacacc	ggaggaatcc	gggggaccca	gatattatgt	gttcgctttg	1080
ctatatgagg	accctgtggca	tgtttgtcta	cagtaaatga	aaattatggg	cagtggggtg	1140
tagagtggtg	ggtttggtgt	ggttaatttt	tttttaattt	ttacagtttt	gtggttttaa	1200
gaattttgta	ttgtgatttt	tttaaaaggt	ccgtgtctcg	aaacctgagc	tgagcccgag	1260
ccagaacccg	agcctgtcca	acctaacccg	cgtcctaaaa	tgggcgctcg	tatcctgaga	1320
gcgccgacat	caactgtgtc	tagagaatgc	aatagtagta	cggatagctg	tcactccggt	1380
ctcttcaaca	caacctcctg	gatacacccc	gtggctcccg	tgtagcccat	taaacacagt	1440
gcccttgagag	ttggtggggc	tcgcacagct	gtggaaatga	tcgagagact	gcttaacagag	1500
ctggggcaac	ctgttgactt	gagctgtaaa	gcccccaggc	cataaggtgt	aaacctgtga	1560
ttgcgtgtgt	ggttaacggc	tttgttgtct	gaatgagttg	atgtaagttt	aataaagggt	1620
gagataatgt	ttaaccttga	tgccgtgtta	aatggggcgg	ggcttaaggg	gtatataagt	1680
cgcgcgtggc	taatttttgt	tacatctgac	ctcatggagc	ctggggagtg	tttgggaagt	1740
ttttctgctg	tgctgaactt	gctgggaacg	agctctaaca	gtacctcttg	gttttggagg	1800

ttctctgtggg	gctcatccca	ggcaaaagtt	gtctgcagaa	ttaaaggagg	ttacaagatgg	1860
gaatttgaag	agctcttggaa	atccctgtggg	gagctgtttg	attcttttgaa	tctgggtgac	1920
caggcgcttt	tccaagagaa	ggctcatcaag	actttggatt	ttccacacc	ggggcgccgt	1980
cgcgctgctg	ttgtcttttt	gagttttata	aaggataaat	ggagcgcaaga	aaccocatctg	2040
agcggggggt	acctgtctgga	ttttctggcc	atgcatctgt	ggagagcggt	tgtgagacac	2100
aaagaatcgcc	tgtacttgtt	gtcttccctg	cgcccgcgga	taataccagc	ggagggagcag	2160
cagcagcagc	aggaggaagc	caggcgccgg	cgccaggagg	agagcccatg	gaaccgcaga	2220
gcgcgccttg	accctcgggg	atgaattgtg	tacaggtggc	tgaactgtat	ccagaactga	2280
gcgcattttt	gacaattaca	gaggatgggc	aggggctaaa	gggggttaag	aggagcgggg	2340
gggcttgtga	ggctacagag	gaggctaggg	atctagcctt	tagcttaatg	accagacacc	2400
gtcctgagtg	tattactttt	caacagatca	aggataattg	gcctaagtga	cttgatctgc	2460
tggcgcaaga	gtattccata	gagcagctga	ccacttaact	gctgcagcca	ggggatgatt	2520
ttgaggaggc	tattagggtg	tatgcaaaag	tggcaacttg	gccagattgc	aagtacaaga	2580
tcagcaaatc	tgtaaatctc	aggaatttgt	gctacatttc	tgggaaacggg	gcgcaggttg	2640
agatagatac	ggaggatagg	gtggccttta	gatgtagcat	gataaatatg	tggcgggggg	2700
tgccttgcat	ggacgggggt	gttatatatg	atgtaaggtt	tactggcccc	aattttagcg	2760
gtacggtttt	cctggccaat	accaacctta	tcctacacgg	tgtaaagctc	tatgggttta	2820
acataacctg	tgtggaagcc	tggacccgat	taagggttcg	gggctgtgcc	ttttactgct	2880
ctgtgaaagg	ggtgggtgtg	cgccccaaaa	gcagggcttc	aattaagaaa	tgccctctttg	2940
aaaggtgtac	ttgtgggttc	ctgtctgagg	gtaactccag	gggtgcgccac	aatgtggcct	3000
cgacgtgtgg	tgtctttcat	ctagtgaaaa	gcgtggctgt	gattaagcat	aacatggtat	3060
gtgcccactg	cgaggacagc	gccttcagga	tgtgcactgc	ctcgccagcg	aactgtccac	3120
tgtcgaagac	cattccagta	gccagccact	ctcgcaagcg	gggggtgttc	cttagcatab	3180
acatactgac	cgctgtttcc	tgtcatttgg	gtaacacagg	gggggtgttc	ctaccttacc	3240
aatgcaattt	gagtcacact	aagatattgc	ttgagcccg	gagcatgtcc	aagtgtaacc	3300
tgaccagggg	gtttgacatg	accatgaaga	tctggaaggt	gctgaggtac	gatgagaccc	3360
gcaccagggt	cagaccctgc	gagtggtggc	gtaaacatat	taggaaccag	cctgtgatgc	3420
tggatgtcag	cgatgaagat	acagattgag	gtactgaaat	gtgtggcggt	ggcttaaggg	3480
tgggaagaaa	tatatgaagt	gggggtctta	tgtagttttg	tatctgtttt	gcagcagccg	3540
cgccgcacct	gagcaccaac	tcgtttgatg	gaagcattgt	gagctcatat	ttgcaaacgg	3600
cgatgcccc	atgggcgggg	gtgcgtcaga	atgtgatggg	ctccagcat	gatggtcgcc	3660
cgctctcgc	cgcacaactc	actaccttga	ctacagagac	cggtgtctga	acgcgctttg	3720
agactgcagc	ctcgcgcggc	gcttcagccg	ctgcagccac	cgcccgcggg	atgtgtgactg	3780
actttgcttt	ctcgtagccg	cttgcaagca	gtgcagcttc	cggttcatcc	gccccgcgatg	3840
acaaagttag	ggctcttttt	gcacaattgg	attctttgac	cggggaaact	aatgtcgttt	3900
ctcagcagct	gttggtatctg	cgccagcagg	ttctgcctct	gaaggcttcc	tcctctccca	3960
atcgcgcttta	aaacataaat	aaaaaacagg	actctgtttg	gatttggatc	aagcaagtgt	4020
cttctgtctc	ttattctagg	gtttttcgcg	cgcggtagcg	ccgggaccag	cgctctcggt	4080
acttgagggt	actgtgtatt	ttttccagga	cggtgtaagg	gtgactctgg	atgttccaat	4140
gggtggcggt	aagccgctct	ctgggggtga	agggagcgctg	ggcggtgggtg	ctaaaaaatg	4200
ctttcagtag	caagctgatt	gccaggggca	ggcccttggt	tgaagttttt	acaaagcggt	4260
taagctggga	tgggtgcata	ctgggggata	tgagatcgat	cttggactgt	atttttagg	4320
tggcttatgt	cccgacata	tcctcccggg	gattcatgtt	gtgcagaacc	accagcaacg	4380
tgtatccggg	gcacttggga	aattttgcac	gtagctcaga	aggaatatcg	tgggaagaact	4440
gtgagacggc	ctgttgacct	ccaagatttt	ccatgcatct	gtccataatg	atggcaatgg	4500
gccacggggc	cgcgccctgg	gcgaagatat	ttctgggata	actaacgtca	tatttgttgt	4560
ccaggatgag	atcgctatag	gccattttta	caaagcgcg	ggggaggggtg	ccagactctgc	4620
gtataatggt	tcacatccgc	ccagggcggt	agttaccctc	acagatttgc	atttccccag	4680
ctttgtggtc	agatgggggg	atcagtctcta	cctgcccggg	gatgaagaaa	acggtttccg	4740
gggtaggggga	gacatcgctg	gaagaaagca	ggttcccgag	cagctgcgac	ttaccgcagc	4800
cggtggggcc	gtataatcaca	cctataccgc	gggtgcaactg	gtagttaaga	gagctgcagc	4860
gtccgtctatc	ctcagcaggg	ggggccactt	cgtttaagcat	gtccctgact	cgcatgtttt	4920
cctctgaccaa	atccgcagga	aggcgctcgc	cgccacgcga	tagcagttct	tgcaggaaag	4980

caaatgtttt	caacgggttg	agaccgtccg	cogtaggcac	gcttttgagc	gtttgaccaa	5160
gcagttccag	cggtgtccac	agctcgggtca	cctgctctac	ggcactcga	tccagcatat	5220
ctccctcggtt	cgcggtgttg	ggcggttttc	gctgtacggc	agtagtcggt	gctcgtccag	5280
acgggcagg	gtcaggtctt	tccacggggc	cagggttctc	gtacgctag	tctgggtcac	5340
ggtgaagggg	tgcgtccgg	gctgcgcgt	ggccagggtg	cgcttagagg	tggctctgct	5400
ggtgctgaag	cgcttcgggt	cttgcgcctg	cgctcgccgc	aggtagcatt	tgacctaggt	5460
gtccatagtc	agccctccg	cggtgtggcc	cttggcgccg	agcttgccct	tggagcaggc	5520
gcgcgcagag	gggcagtgca	gactttttag	ggcgttagag	ttggggcgca	gaataaccga	5580
ttccggggag	taggcacatc	cgccgcaggc	cccgacagac	ccatgctttt	ccacagacca	5640
ggtgagctct	ggcgtgtcgg	gggtcaaaaa	cagggttccc	cgccgttctt	tgatcggttt	5700
cttacctctg	gtttccatga	gcccgtgtcc	agctcgggtg	acgaaaaagg	tgctccgtgt	5760
cccgatatca	gacttgagag	gctcgtctct	gagcgtgttt	cgcgggtcct	ccctgtatag	5820
aaactcggac	cactctgaga	caaaaggtcg	cgctccaggc	acgcagaagg	aggctaagtg	5880
ggaggggtag	cggtcgttgt	ccactagggg	gtccactcgc	tccagggtgt	gaagacacat	5940
gtcgccctct	tcggcatcaa	ggaaggtgat	tgggttttag	gtgtaggcca	cgtagccggg	6000
gtttcctgaa	ggggggctat	aaaagggggt	ggggggcgcg	tgcctctcac	tctcttcgcg	6060
atcgctctct	gcgagggcca	gctgttgggg	tgagtactcc	ctctgaaaag	cggtcatgac	6120
ttctgcgcta	agattgtcag	tttccaaaaa	cgaggaggat	tgtgatatta	cttggccgcg	6180
ggttagtgcct	ttgaagggtg	cgcatcccat	ctggtcagaa	aagacaatct	ttttgttgtc	6240
aaagtgtggt	gcacaaagacc	cgttagagggc	gttggacagc	aacttggcga	tggagcgcag	6300
ggtttgggtt	tgtgcgcgat	cgccgcgctc	cttggccgcg	atgttttagct	gcacgatttc	6360
gcgcgcacac	caccgccatt	cggtgaaagac	ggtggtgcgc	tctgcggcca	ccaggtgcaac	6420
gcgcacacac	cggttgtgca	gggtgacaaag	gtcaacgctg	gtggtctact	ctcccggtag	6480
gcgctcgttt	gtccagcaga	ggcgccgcgc	cttgcgcgag	cagaatggcg	gtagggggctg	6540
tagctgcgct	tgcctcgggg	ggtctgcgtc	cacggtaaaag	accccgggca	cgagcgcgcg	6600
gtcgaagtga	tctatcttgc	atccttgcaa	gtctagccgc	tgtcgcacat	cgccggcgccg	6660
aagcgcgcgc	tctatgttgt	ttagtggggg	accccatggc	atgggggtgg	tgagcgcgga	6720
ggcgtacatg	ccgcaaatgt	cgtaaacgta	gaggggctct	ctgagtattc	caagatatgt	6780
agggtagcat	cttccaccgc	ggatgctggc	gcgcacgtaa	tctgtatagt	cgtagcaggg	6840
acgcagagag	tcgggaccca	ggttgctacg	ggcggtctgc	tctgctcgga	agactatctg	6900
cctgaagatg	gcattgtagt	tggatgatat	ggttggaagc	tggaaagacgt	tgaagctggc	6960
gtctgtgaga	cctaccgcgt	cacgcacgaa	ggagcgctag	gagtcgcgca	gcttgttgac	7020
cagctcggcg	gtgacctgca	cgcttagggc	gcagtagtcc	agggtttcct	tgatgatgtc	7080
atacttatcc	tgtccctttt	tttccacag	ctcgcggttg	aggacaaact	cttcgcggtc	7140
tttccagtat	tctttgtagc	gaaacccgtc	ggcctccgaa	cggtaaagac	ctagcatgta	7200
gaactgtgtt	agcgcctggt	agcgcacgca	tcccttttct	acgggttagc	cgtagctgct	7260
cgccgctctc	cgagcggagg	tgtgggtgag	cgcaagggtg	tccctgacca	tgaactttgag	7320
gtactggtat	ttgaagtgcg	tgtcgtcgca	tccgccctgc	tccacagcca	aaaagtcctg	7380
gcgctttttg	gaacccgagt	ttggcagggc	gaaggtgaca	tcttgaaga	gtatcttttc	7440
cgccgcagga	ataaagtgtc	gtgtgatgcg	gaagggtccc	ggcaccctcg	aacggttgtt	7500
aattacctgg	gcggcgagca	cgatctcgtc	aaagcgtgtg	atgtttgtgc	ccacaatgta	7560
aaagtccaag	aagcgcggga	tgcctctgat	ggaaggcaat	tttttaagtt	cctcgttagt	7620
gagctctcca	ggggagctga	gcccggtgtc	tgaaggcccc	cagtttgcaa	gatgagggtt	7680
ggaagcgagc	aatgagctcc	acaggtcacg	ggccatttag	atttgcaggt	ggtcgcgaaa	7740
ggctcctaac	tggcgacctc	tggccatttt	ttctgggggtg	atcagtaga	aggttaagcg	7800
gtcttgtctc	cagcgttccc	atccaaaggt	cgcgcttagg	tctcgccggg	cagtcactag	7860
agggtcatct	ccgcgcgaat	tcatgaccac	catgaagggc	acagcgtgct	tcccaaaagg	7920
ccccatccaa	gtataggtct	ctacatcgta	ggtgacaaag	agacgctcgg	tgcgagggat	7980
cgagcgcgtc	gggaagaaact	ggatctcccg	ccaccaattg	gaggagtggc	tattgatgtg	8040
gtgaagtagc	aggtccctgc	gacgggcgca	acactcgtgc	tggcttttgt	aaaaaagtgc	8100
cgagtactgg	cagcgggtgca	cggtctgtac	atctctgca	aggttgacct	gacgcacccg	8160
caacaggaag	cagagtggga	atttgagccc	ctcgcctggc	gggtttggct	gggtgtcttc	8220
tactctggct	gcttgtctct	gaccgtctg	ctgctcgaag	ggagttacgg	tggatcgga	8280
caccacggcg	cgcgagccca	aagtccagat	gtccgcgcgc	ggcgttcgga	gcttgatgac	8340
aacatcgccg	agatggggag	tgtccatggt	ctggagctcc	cgccgctcta	ggtcaggcgg	8400

gagctcctgc	aggtttacct	cgcataagacg	ggtagcggcg	cgggctagat	ccagggtgata	8460
cccaatttcc	agggctcggt	tggtagcggcg	gtcgaatggct	tgcgaagaggc	cgcattccccc	8520
cggcgcgact	cggctaccgc	cggcgcgggcg	gtggcgcgcg	gggggtgtcct	tggatgatgc	8580
atctaaaaagc	ggtagcggcg	ggagaccccc	ggaggtaggg	ggggctccgg	accgcgcggg	8640
agaggggggga	ggggcagctc	ggcgccggcg	cgggcgaggga	ctgggtgtctg	cgcgcgtatgg	8700
ttgtcggcgga	cgcgcagcagc	cggcgcggttg	atctcctgtaa	tctggcgccct	ctcgctgaaag	8760
acgcagcggcc	cggtagcgtt	gagcctgaaa	gagagttcga	cagaatacaat	tccgggtgtcg	8820
ttgacgcggcg	ctcggcgcaa	aatctcctgc	acgtctcctg	agttgtctctg	ataggcgatc	8880
tcggccatga	actgcctgat	ctctcctccc	tggagatctc	cgcgtccggc	tcgctccacg	8940
gtggcggcgga	ggtagcttggg	aatcggggcc	atgagctgcg	agaaggcggt	gaggcctccc	9000
tcgttccaga	cgcggctgta	gaccacgccc	ccttcggcat	cgcggggcgcg	catgaccacc	9060
tgcgcagat	tgaactccac	gtgcgcggcg	aagacgcggt	agtttcgcag	cgctgaaag	9120
aggtagttga	ggtaggtggc	ggtaggttct	gccacgaaga	agatcataaac	ccagcgtcgc	9180
aacgtggatt	cgttgtatata	ccccaaggcc	tcaaggcgct	ccatggcctc	gtagaagtcc	9240
acggcgaaagt	tgaaaaaactg	ggagtgcgc	gcgcacacgg	ttaactctct	ctccagaaga	9300
cggatgagct	cggcgacagt	gtcgcgcacc	tcgcgtctaa	aggctacagg	ggcctctctt	9360
tctcttccaa	tctctctctc	ataaaggccc	tccctctctt	ctctctctgg	cggcggttggg	9420
ggaggggggga	cacgcggggcg	acgacggcg	acggggaagg	ggctcgacaaa	cgctctcgatc	9480
atctccccgc	ggcgcagcgc	catggtctcg	gtgacggcg	ggccgtctctc	gcggggggcgc	9540
agbtggaaga	cgcgcggcgt	catgtcccgg	ttatgggttg	gcgggggggtg	gccatcgggc	9600
agggatacgg	gcataacagt	gcattctaac	aattgttgtg	taggtactcc	gcgcgcgaag	9660
gacctgagcg	agtcgcgcatc	gaccggatcg	gaaacacctc	cgagaaaggc	gtctaacaccg	9720
tcacagctgc	gaatgagctg	gagcaccggt	gcggggcgcca	gcgggcggcg	gtccgggttgc	9780
ttcttgcggcg	aggtgtcgtc	gatgatgtaa	ttaaggttagg	cggctcttag	acggcggaatg	9840
gtcgacagaa	gcacacatgc	cttgggtccg	gcctgtcgaa	tgcgcaggcg	gtcggccatg	9900
ccccaggett	cgttttgaca	tcggcgcgag	tctttgtagt	agtccttgcat	gagccttctt	9960
acgggcaett	cttctctctc	tctctcttgt	cctgcattat	ttgcattctat	cgtcggcgcg	10020
cgggcgaggt	ttggccgtag	gtggcgccct	cttctctcca	tgcgtgtgac	cccgaagccc	10080
ctcatcgccct	gaagcagggc	taggtcggcg	acaacgcgct	cggctaata	ggcctgtctg	10140
acctcgctga	gggtgagact	gaagtcatcc	atgtccacaa	agcggtagta	tgcgcccggt	10200
ttgatgggtg	aagtgcagtt	ggccataaac	gaccagttaa	cggctctgggt	accgcgctgc	10260
gagagctcgg	gttacctgag	acgcgagtaa	gcctctcagt	caaatacgtg	gtcgtttgcaa	10320
gtccgcacca	ggtactggta	tcccacccaa	aaagtgcggc	gcggctggcg	gtagagggggc	10380
cagcgtaggcg	tggccggggc	tccggggggc	agattcttcca	acataaaggc	atgataatccg	10440
tagatgtaac	tggacatcca	ggtagtcggc	cgggcggtgg	tggaggcgcg	cggaaagtgc	10500
cgagcgcgggt	tcgcagtggt	gcgcagcggc	aaaaagtgtc	ccatggctcg	gacgtcttgg	10560
cgggttaggc	gctgcgaatc	gttgacgctc	tagaacgtgc	aaaaggagag	cagcttaagcg	10620
ggacattctc	cgtgtctctg	tggataaat	cgcaagggtg	tcatggcgga	cgaccgggggt	10680
tcgagccccc	tatccggcgc	tccgcggtga	tccatccggt	taccgcccgc	gtgtcgaac	10740
caggtgtcgc	acgtgacaga	acgggggagt	gctccttttg	gctctcttcc	agggcgggcg	10800
gctctgcgc	tactgttttt	ggccactggc	cgccgcagc	gtaagcgggt	agcctggaaa	10860
cggaagaact	taagtgtctc	gctcccgtga	gcggaggggt	tattttccaa	gggttgagtc	10920
cggggacccc	gggttcgagt	ctcggacggc	ccggaactgc	gcgaacgggg	gtttgctctc	10980
cgtcatcgca	agaccgccct	tgcataatcc	tccggaaaca	gggacgagcc	ccttttttgc	11040
tttttccaga	tgcattccgt	gctcgggcag	atggccccc	ctcttcagca	gcggcaagag	11100
caagagcagc	ggcagacatg	caggggaccc	tccctctctc	ctaccgcgct	aggaggggcg	11160
acatcccggt	ttgacgcggc	agcagatggt	gattacgaac	ccccggcgcg	ccggggccggg	11220
caactacctg	acttggagga	gggcgagggc	ctggcccgcg	taggagcgcc	ctctctctag	11280
cggtagccaaa	gggtgcaagt	gaagcgtgat	acgcgtgagg	cgtactgccc	ggggcgaaac	11340
ctgtttccgc	accgcgaggg	agagggagcc	gagagagatg	gggatcgaaa	gttccacgca	11400
gggcgcgagc	tgcggcatgg	cctgaatcgc	gagcggttgc	tgcgcgagga	ggactttbag	11460
ccgcagcgcg	gaaccgggat	tagtcccgcg	cgcgcacacg	tgcggccgcg	gcaccttgta	11520
accgcatacg	agcagcaggt	gaaccaggag	attaaatttc	aaaaagactt	taacaaaccac	11580
gtgcgtacgc	tgtgtgcgcg	cgaaggagtg	gctataggag	tgtatgcatt	gtgggaacttt	11640
gtaagcgcgc	tggagcaaaa	cccaaatagc	aagccgctca	tggcgacgct	gtctcttata	11700

gtgcagcaca	gcaggagacaa	cgaggcattc	agggatgcgc	tgctaaacat	agtagagccc	11760
gagggcgctg	ggctgctcga	tttgataaac	atccctgcaga	gcatactggg	gcaggagcgc	11820
agcttgtagct	tggtcgacaaa	ggctggccgcc	atcaactatt	ccatgcttag	cctggggcag	11880
ttttacggcc	caagagatata	ccataccccc	tacgttccca	tagacaagga	ggtaaaagtc	11940
gagggggctct	acatgcgcact	ggcgctgaag	gtgcttacct	tgagcgcaga	cctgggctgt	12000
tatcgcaaga	agcgcatcca	caaggccgtg	agcgtgagcc	ggcgccgcga	gctcagcgac	12060
ccgcgactga	tgacacgect	gcaaaaggccc	ctggctggca	cgggcagcgg	cgatagagag	12120
gcccagctct	acttttgagc	ggcgctgac	ctgcgctggg	ccccaaagcc	acgcgcctcg	12180
gaggcagctg	ggggccggacc	tgggctggcg	gtggcaccgc	cgcgcgctgg	caacgtcggc	12240
ggcgctggagg	aatatgacga	ggacgatgag	tacgagccag	agggacggcg	gagctaacgc	12300
gtgatgtttc	tgatcagatg	atgcaagacg	caacggaccc	ggcggtgcgg	gcccgcctgc	12360
agagggcagcc	gtccggcctt	aactccacgg	acgactggcg	ccaggtcatg	gaccgcatac	12420
tgtcgctgac	tgcgcgcaat	cctgacgcgt	tcgggcagca	gcgcgagccc	aaccggctct	12480
ccgcgaattct	ggaagcgggt	gtccccggcg	gcgcgaaccc	cacgacagag	aaggtgctgg	12540
cgatcgtaaa	cgcgctggcc	gaaaacacgg	ccatccggcc	cgacgagggc	ggcctggctt	12600
acgacgcgct	gcttcacgcg	gtggctcgtt	acaacagcgg	caacgtgcag	accacactgg	12660
acggcgctgt	ggggctagtgt	cgcgaggcgg	tgggcgagcg	tgagcgcgcg	cagcagcagg	12720
gcaacactggg	ctccatgggt	gcactaaacg	cttctctgag	tacacagccc	gccaaactga	12780
cgcgggggaca	gggagactac	accacacttg	tgagcgcaat	cgggctaatg	gtgactgacg	12840
caccgcgaag	tgaggtgtac	cagctctggg	cagactattt	tttccagacc	agtagacaga	12900
gcctgcagac	gcttaaacctg	agccaggctt	tcaaaaactt	gcaggggctg	tggggggtgc	12960
ggggtccccc	agggcagccg	gcgacactgt	ctagcttgct	gagcggcaac	tcgcgcctgt	13020
tgctgtgctc	aattgagccc	ttccagggca	gtggcagcgt	gtccccggag	acacacttag	13080
gtcaactgct	gacactgtac	cgcgaggcca	taggtcaggc	gcattgtggc	gagcatcact	13140
tcacagagat	tacaagtgtc	agccgcgcgc	tgggggcagc	ggacacgggc	agcctggagg	13200
caaccctaaa	ctacctgctg	accacacggc	ggcagaagat	ccctcgttgc	cacagttaa	13260
acagcagagg	ggagcgcatt	ttgcgctacg	tgacgacag	cgtgagcctt	aacctgatgc	13320
gcgacggggt	gaagcccgag	gtggcgctgg	acatgacccg	gcgcaacatg	gaaccgggca	13380
tgatgccttc	aaacggcgcg	tttatcaacc	gcctaatgga	ctacttgcat	cgcgcgggcg	13440
ccgtgaaccc	cagtagtatt	accaatgcca	tcttgaaccc	gcactggcta	cgccccctgt	13500
gtttctacac	cgggggattc	gagggtgccg	agggtaacga	tggaattcct	tgggacgaca	13560
tagacgacag	cgtgttttcc	ccgcaacccg	agaccctgct	agagttgcaa	cagcgcgacg	13620
agggcagggc	cgctctgcga	aaggaaagct	tcgcagggcc	aagcagcttg	tcgcatctag	13680
gcgctgcggc	cccgcggtca	gatgctagta	gcccatttcc	aagcttgata	gggtctctta	13740
ccagactctg	caccacccgc	ccgcgcctgc	tgggcgaggga	ggagtagcta	aaacaactgc	13800
tgctgcagcc	cgagcgcgaa	aaaaacctgc	ctccggcatt	tcccaacac	gggtagagga	13860
gcttagtgga	caagtagagt	agatgggaag	cgtagcgcca	ggagcacagg	gagctgccag	13920
ccccgcggcc	gccaccacct	cgtcaaagcg	acgcagctca	cgggggctcg	gtgtggggag	13980
acgatgactc	ggcagacgac	agcagcgtcc	tggaattggg	agggagtgcc	aaccggtttg	14040
cgaccccttc	ccccagcgtg	gggagaaatg	tttaaaaaaa	aaaaagcatt	atgcaaaata	14100
aaaaactcac	aaagggccatg	gcaccgagcg	ttggttttct	tgatttcccc	ttagtatgcg	14160
ggcgcgggcg	atgtatgagg	aaggtctctc	tcctctctac	gagagtggtg	tgagcgccgc	14220
ggcagtgagg	ggcgcgctgg	gttctctcct	cgatgctccc	ctggaccgcg	cgtttgtgtc	14280
tcggcggtac	ctgcggccta	ccggggggag	aaacagcacc	cgttactctg	agttggcacc	14340
ccattatcgac	accaccggcg	tgtagcttgt	ggacaacaag	tcacgggatg	tgcatccctt	14400
gaactaccag	aagaccacca	gcaactttct	gaccacggct	attcaaaaca	atgactacag	14460
ccggggggag	gcaagacacac	agaccatcaa	tcttgacgac	ggggcgggca	cgggcgggca	14520
cctgaaaaaa	atcctgcata	ccaacatgcc	aaatgtgaac	gagttcatgt	ttaccaataa	14580
gtttatggcg	cggttgatgtg	tgtcgcgctt	gcttactaag	gacaaatcag	tggaagctga	14640
atacgagtgg	gtggagttca	cgtgcgcgga	gggcaactac	ggcagacaga	acgggggtct	14700
ccattatgaac	aacgcgatag	tggaagcact	cttgaaagtg	ggcagacaga	acgggggtct	14760
ggaagcgcac	atcggggtga	agtttgacac	ccgcaacttc	ttccatccag	acatcatttt	14820
cactgttcctt	gtctatgcctg	gggtatatac	aaacgaagcc	agcaacttgt	tgggcatccg	14880
gtcgcaggga	tgcggggtgg	acttcaacca	cagccgctgt	gatgatctgg	agggtagtaa	14940
caagcgccaa	cccttcaggg	agggctttag	gatcacctac			15000

cattcccgcga	ctgtttggatg	tggagccta	ccaggcgagc	ttgaagagatg	acaccgaaca	15060
ggcggggggt	ggcgagggcg	gcagcaacag	cagtgggcagc	ggcggggaag	agaactccaa	15120
cgccgagcgc	ggggcaatgc	agccgggtgga	ggacatgaac	gatcatggcca	tcgcggggga	15180
cacctttggcc	acagcggcgct	aggagagaagc	cgctgaggccg	gaagcagcggc	ccgaagctgc	15240
gcctcccgct	gcgcaaccgc	aggtcgagaa	gcttcagaag	aaaccgggtga	tcaaacccct	15300
gacagaggagc	agcaagaaac	gcagttacaa	cctaataaagc	aatgacagca	ccttcaccga	15360
gtacacgcag	tggttaccctg	catcaacaata	cggcgaccctc	cagaccggaa	tcgcgtcatg	15420
gacccctgctt	tgcaactcctg	acgtaaacctg	cggctcggagc	caggtctact	ggctcgttgc	15480
agacatgatg	caagaccccg	tgacottcgc	ctccacgcgc	cagatcagca	actttccggt	15540
ggtggggcgcc	gagctgttgc	ccgtgcactc	caagagcttc	tacaacgacc	aggccgtcta	15600
ctcccaactc	atccgccagt	ttacctctct	gacccaagtg	gtcaatcgct	ttcccgagaa	15660
ccagattgtt	ggcgcccgcc	cagcccccac	catcaccacc	gtcagtgaaa	acgttccctg	15720
tctcacagat	cagggagcgc	taccgtctgc	caacagcattc	ggaggagtc	cgcgagtgac	15780
catattactga	gcagagcgc	gcaccttgcc	ctacgtttac	aaggccctgg	gcattagtctc	15840
gcgcgcgcgtc	ctatcgagcc	gcactttttg	agcaagcatg	tccatcctta	tatcgccag	15900
caataacaca	ggctggggccg	tgccttcc	aagcaagatg	tttggcgggg	ccaagaagcg	15960
ctccgaccac	caccagatgc	gcgtgcgcgg	gcactaccgc	gcgcctggg	gcgcgcacaa	16020
ccgcggccgcg	actgggcgca	ccaccgtcga	tgacgccatc	gacgcgggtg	tgaggaggcg	16080
gcgcaactac	acgcccaagc	cgccaccagt	gtccacagtg	gacgcggcca	ctcagaccgt	16140
ggtygcgcgga	gcccgccgct	atgtctaaat	gaagagacgg	cggaggcgcg	tagcacgctgc	16200
ccatccgcgcg	gcacccggca	ctgcgcgcga	acgcgcggcg	cgccgccctgc	ttaccgcgcg	16260
acgtcgcacc	gcgcgcggg	cggccatgcg	ggccgctcga	aggctggcg	cgggtatgtg	16320
cactgtgccc	cccgagttcca	ggcgacgagc	ggccgcgcga	gacgcgcgg	ccattagtcg	16380
tatgactcag	ggtcgcaggg	gcaacgtgta	ttgggtgcgc	gactcgggtta	gcggccctgcg	16440
cgtygcgcgtg	ctgcaccgcc	ccccgcgcaa	ctagattgca	agaaaaaact	acttagactc	16500
gtactgttgt	atgtatccag	cggcgccgcg	ggcgaacgaa	gctatgtcca	agcgcaaaat	16560
caaagaagag	atgctccagg	tcatcgccgc	ggagatctat	ggccccccga	agaagggaaga	16620
cgaggattac	aagcccgcaa	agctaaagcg	ggtcaaaaag	aaaaagaag	atgatgatga	16680
tgaacttgac	gacgaggtgg	aactgctgca	cgctaccgcg	cccaggcgac	gggttacagt	16740
gaaaggtcga	cgcgtaaaac	gtgttttgcc	accgcgcacc	accgtatgct	ttacgcgcgc	16800
tgagcgctcc	accgcacact	acaagcgcgt	gtatgatgag	gtgtacggcg	acgaggacct	16860
gcttgagcag	gcacacgagc	gctcggggga	gtttgctact	ggaaagcgcc	ataaggacat	16920
gctggcggttg	cgcctyggacg	agggcaaccc	aacacctagc	ctaaaagccg	taacactgca	16980
cgaggtgctg	cccgcgcttg	caccgtccga	agaaaagcgc	ggcctaagcg	gcgagttcgg	17040
tgactttggca	cccaccgtgc	agctgatggt	acccaagcgc	cagcgactgg	aaagtatgctt	17100
ggaaaaaatg	accgtggaac	ctgggctgga	gcccaggttc	cgctgcggcg	caatcaagca	17160
ggtggcgccg	gactgtggcg	tgacagccgt	ggacgtttac	ataccacta	ccagtagcac	17220
cagttatgcc	accgccacag	agggcattga	gacacaaagc	tcccgggttg	cctcagcggt	17280
ggcggtatgc	gcggtgcagg	cgttcgctgc	ggccgcgttc	agaactctca	cggaggtgca	17340
aaagccagccg	tgagatttgc	gcgtttcagc	ccccgcgcgc	ccgcgcggtt	cgaggaagta	17400
cggcgccgcg	agcgcgctac	tgccccgaata	gcctctacat	ccttcattg	tgccaccoc	17460
cggctatcgt	ggctacactc	accgccccag	aaagcagaca	actaccgcac	gcgcgaacac	17520
cactggaacc	cgcgcgcgc	ctgcgcgtgc	ccagcccgct	gtccgcggca	ttcgctgcg	17580
caggggtgct	cgcgaagagc	gcaggaccct	ggtgtgcga	acagcgcgct	acaccccgac	17640
catcgtttaa	aagccggctc	ttgtggttct	tgcatgatg	gcctcactc	gcgcgctccg	17700
ttcccggtg	cggggaattcc	gagggaagaat	gcaccgtagg	agggggcatg	ccggccacgg	17760
cctcagcgccg	ggcatgtctc	gtgcgcacca	ccggcgccgg	cgccgcctgc	accgtcgcat	17820
gcgcgpcgggt	atctcgcc	tccttattcc	actgatccgc	cgccgattg	ggccgctgoc	17880
cggatattga	tcgctggcct	tgccggcgca	gagacacttg	ttaaaaacaa	gttgcatgtg	17940
gaaaaatcaa	aataaaaaatg	ctggaactctc	acgcctcgctc	gttcctgttaa	ctatttttga	18000
gaatggaga	catcaacttt	gcgtctctgg	ccccgcgaca	cgcctcgcgc	cggttcatg	18060
gaactcgcca	agatattcggc	accagcaata	tgagcggtgg	gccttcagc	tgggcctcgc	18120
gtggagcgcg	catcaaaaaat	cttggtttcca	ccgttaagaa	ctatggcagc	aaggccctga	18180
acagcagcac	agccagcatg	ctaggggata	agttgaagaa	gcaaaatttc	caacacagac	18240
tgttagatgg	cctggccctc	ggcattagcg	gggtggtgga	cctggccaac	caggcagtcg	18300

aaaataagat	taacagtaag	cttgatcccc	gccctcccg	agaggagcct	ccaccggccg	18360
tgagagagac	gtctccagag	ggcgctggcg	aaaagcgctc	gcgcccgcac	agggaagaaa	18420
ctctggtagac	gcaaatagac	gagcctccct	cgtagcagga	ggcactaaag	caagcgctgc	18480
ccaccacccg	tcctatcgcg	cccatggcta	cggagtgct	ggccagcac	acaccgctaa	18540
cgcctggacc	gectcccccc	ggcgacaccc	agcagaaccc	tgtgctgcc	ggccgagccg	18600
cogtctggtt	aaacccgtct	agccgcgcgt	ccctgcgcgc	cgccgcagc	ggctcgcgat	18660
cgttgcgccc	cgttagccag	ggcaactggc	aaagcacact	gaacagcact	gtgggtctgc	18720
gggtgcaatc	cctgaagcgc	cgcagatgct	ctcgaatgac	taacgtgtgc	tatgtgtgtc	18780
atgtatgctg	ccatgtcgcc	ggcagaggag	ctgctgagcc	ggcgcgccgc	cgctttccaa	18840
gatggotacc	ccttgatgat	tgccgcagtg	gtcttcatat	caatctcgg	cgtaactcag	18900
ctcggagtag	ctgagccccc	ggctgggtgc	gtttgcgcgc	ggcaccgaga	ccacagagcc	18960
ctgtaataac	aagtcttagaa	acccacccgt	ggcgctcag	cagcagctga	cgtaactcga	19020
gtcccagcgt	ttgacgctgc	ggcttcaccc	tgtggaccgt	gaggaacttg	cttccacgta	19080
caagggcgcg	ttcaccctag	ctgtgggtga	taacgtgtgc	ctggacatgg	cttggcactgc	19140
ctttgacatc	cgcggcgctgc	tggaacaggg	ccctactttt	gaatgggatg	aagctgctac	19200
ctacacgcgc	ctggctccca	agggtgcccc	aaatccttgc	gaagaagga	taaacattac	19260
tgctcttgaa	ataaacctag	aagaagagga	cgatgacac	tattctggta	taaaatacct	19320
agctggagac	caaaaaactc	acgtattttg	gcaggcgctc	taaatcacct	ataaaacatt	19380
aaaggggggt	attcaaatag	gtgtogaagg	tcacacacct	aaatatgcgc	ataaaacatt	19440
ctaacctgaa	cctcaaatag	gagaatctca	gtggtacgaa	actgaatata	atcatcgagc	19500
tggaagagtc	cttaaaaaag	ctaccccaat	gaacacatgt	taagggttat	atgcgaacac	19560
cacaaatgaa	aatggggggc	aaggcatctt	tgtaaagcaa	caaaatggaa	agctagaagag	19620
tcaagtgtaa	attgcaatttt	tctcaactac	tgagggcgac	cgagggcaatg	gtgataactt	19680
gactcctaaa	gtggttatgt	acagtgaaga	tgtagatata	gaaccccgac	acactcatat	19740
ttcttcatcg	cccactatta	aggaaggtaa	ctcacgagaa	ctaatgggcc	aacaaacttat	19800
ggccacacgg	cctaattaca	ttgcttttag	ggacaaatct	attggtctaa	tgtattacaa	19860
cagcacgggt	aatatgggtg	ttctggcgcg	ccaagcatcg	cagttagaatg	ctgtttagta	19920
tttgcaagac	agaacacag	agctttcata	ccagcttttg	cttgattcca	ttggtgatag	19980
aaccaggtac	ttttctatgt	ggaatcaggc	tgttgacagc	tatgatccag	atgttagaat	20040
tattgaaatc	catgggaactg	aagatgaact	tcacaaattc	tgctttccac	tgaggaggtg	20100
gatataatca	gagactctta	ccaaggtaaa	acctaaaaca	ggcaggagaa	attggtggga	20160
aaagatgctc	acagaaatct	cagataaaaa	tgaataaaga	gttggaataa	attttgccat	20220
ggaatcaat	ctaaatgcca	acctgtggag	aaatttctgt	tactocaaca	tagcgctgta	20280
tttgcccgac	agctaaaagt	acagtctctc	caacgtaaaa	atttctgata	accocaaacac	20340
ctacgactac	atgaacaagc	gagtgtgtgc	tcccggtgtc	gtggactgct	acat taacct	20400
tgaggacgcg	tggtccctgt	actatatgga	caacgtcaac	ccatttaacc	accaccgcaa	20460
tgctggcctg	cgtcacgcgt	caatgttgct	gggcaatggt	cgctatgtgc	ctctccacat	20520
ccagggtgct	cagaagtctt	ttgccattaa	aaacctcctt	ctcctgccgc	gctcatcac	20580
ctacgtgtgg	aactccagga	aggatgttaa	catggtttct	cagagctccc	taggaataat	20640
ctctagggtt	caggagacca	gcatataagt	tgatagcatt	tgcccttaac	ccacctctct	20700
ccccatggcc	gaacacacgc	cctccacgct	tgaggccatg	cttagaaacg	acacaaagca	20760
ccagctcttt	aacgactatc	tctccgcgcg	caacatgctc	taccctatac	cggccacgcg	20820
taccacgcctt	cccatatcca	tcacctccgc	caactggggc	gctttccgcg	gttcgcgctt	20880
cacggcgctt	agaagcaagg	aaaccccatc	actggcgctg	ggctacgacc	cttattacac	20940
ctactctggc	tctataccct	acctagatgg	aaacttttac	ctcaaccaca	ctttaagaaa	21000
gggtggcatt	accttgagct	cttctgtcag	ctggcctggc	aatgacccgc	tgcttaaccc	21060
caacgagttt	gaatttaagc	gctcagttga	cggggagggg	tacaacgttg	ccaggtgtaa	21120
catgaccaaa	cagtggtttcc	tggtacaaat	gctagctaac	tacaacatgt	gctaccaggg	21180
ctctctatato	ccagagagct	acaaggacgc	catgactccc	ttctttagaa	acttccagcc	21240
catgagcgtg	caggtgtgtgg	atgatactaa	atacaaggac	ttaccaacgg	tgggcatcct	21300
acacacacac	aaacactctg	gatttgtgtg	ctactctggc	ccccaccatg	ggcagggaca	21360
ggcctaccct	gctaaacttc	ccctaccgct	tataggccag	accgcagttg	cagcatttac	21420
ccagaaaaag	ttcttttggc	atcgacacct	ttggcgcatc	ccattctcca	gtaacttat	21480
gtccatgggc	gctcatcag	acctggcgca	aaacctcttc	taegcaactc	cggcccaagc	21540
gctagacatg	acttttgtgg	tggtaccatc	ggacgagccc	acccttcttt	atgtttgtgt	21600

tgaagtcttt	gaogctgttc	gtgtgcacgc	gccgcacccg	ggcgctcatcg	aaaccgtgta	21660
cctgcgcagc	ccctctctcg	ccgcgaacgc	cacaacataa	agaagcaagc	aacatcaaca	21720
acagctggcg	ccatgggctc	cagtgcagcg	gaactgaagg	ccattgtcaa	agatcttggg	21780
ttggggccat	atttttttgg	caacctatgc	aagcgctttc	caggctttgt	ttctccacac	21840
aagctcgctc	ggcccatagt	caatacggcc	ggtcgcgaga	ctggggcgct	acactgggat	21900
gcctcttgcct	ggaacccgca	ctcaaaaaa	tgctacctct	ttgagccctt	tggtctttct	21960
gcacagcgac	tgaaacaggt	ttaccagttt	gagtagcagc	caactcctgc	cgttagcgcc	22020
attgtctctt	ccccgcagcg	ctgtataacg	ctggaaaagt	ccaccocaa	cgtacagggg	22080
cccaactcgg	ccgcctgttg	actattctgc	tgcatgtttc	ttccgcctt	tgccaactgg	22140
ccccaaactc	ccatggatca	caaccaccac	atgaacctta	ttaccggggt	accocaaactc	22200
atgtctcaaca	gtccccaggt	acagcccacc	ctgcgtcgca	accaggaaca	gctctacagc	22260
ttcttgggagc	gccactcgcc	ctacttccgc	agccacagtg	cgcagatag	gagcgccact	22320
ttctttttgtc	acttgaaaaa	catgtaaaaa	taatgtacta	gagacacttt	caataaaggc	22380
aaatgctctt	atttgtacac	ctcgggtgta	ttatttacc	ccacccttgc	cgtctgcccc	22440
gttataaaat	caaaaggggtt	ctgccgcgca	tcgctatgcg	ccactggcag	ggacacgttg	22500
cgatactggc	gtttagtgtc	ccacttaaac	tcaggacaaa	ccatccgcgg	cagctcggtg	22560
aagtttttca	ttccacaggt	gcgcaccatc	accacacgct	ttagcaggtc	ggggcgccat	22620
actctgaagt	cgagttggg	gcctcccgcc	tgccgcgcgc	agttgcgata	cacaggggtg	22680
cagcaactgga	acactatcac	cgccgggttg	tgcaacgctg	ccagcacgct	cttgtcgag	22740
actcagatct	cgctccaggtc	ctccgcgttg	ctcaggggca	acggagttca	ctttgttagc	22800
tgcttccca	aaaaggggcg	gtgccacagg	ttttagttgc	actcgcacgc	tagtggcctc	22860
aaaagctgac	cgctgcgggt	ctggggcgta	ggatcacagc	ccctgcataa	agccttgatc	22920
tgcttaaaa	ccactcagc	ctttgcgcct	tcagagagaa	acatgccgca	agactttgcc	22980
gaaaaactgat	tgccggcgaca	ggccgcgtgc	tgcaacgcgc	accttgcgtc	gggtttggag	23040
aatgcaccca	catctcgccc	ccaccgggtc	ttccagatct	tgcccttgct	agactgtctc	23100
ttcagcgccg	gttgcgcgtt	ttcgtctctc	acatccattt	caatcacgtg	ctccttattt	23160
atcataatgc	ttccgtgtag	acacttaagc	tcgccttcga	tctcagcgca	cggtgtcgagc	23220
cacacagcgc	agcccggtgg	ctcgtgtatc	ttgtaggtca	ctctgcataa	cgaactcgag	23280
tacgcctgca	ggaatcgccc	catcatctgc	acaaaggtct	tggtgctggt	gaaggtcgag	23340
ttcgaacccg	gggtctcttc	gttcagccag	gtcttgcata	cggccgcca	agcttccact	23400
tggtcgagga	gtagtgtgaa	gttcgccttt	agatcgttat	ccactgtgta	cttgtccatc	23460
agcgcgcgcg	gaactccat	gcccttctcc	acgcagagca	cgatcgccac	actcagcggg	23520
ttcatccacg	taatttcaat	ttccgctctc	ctgggctctt	ctcttctctc	ttgcgtccgc	23580
ataccacccg	ccactgggtc	gtcttctatc	agccgcgcga	ctgtgccttt	acctcttttg	23640
ccatgcttga	ttagcacccg	tggtgtgtgc	aaaccccaac	tttgtagcgc	caacatctct	23700
ctttcttctc	cgctgtccac	gattacctct	ggtgatggcg	ggcgctcggg	cttgggagaa	23760
ggcgctcttg	tttttctctt	ggcgccaatg	gccaaatccg	ccgcggaggt	cgtgtgcgcg	23820
ggcgctgggtg	tgccgcgcac	cagcgcgtct	tgtgatgagt	cttctctctc	ctcggactct	23880
atagccgcgc	tcacgcgttt	ttttgggggc	gccggggag	ggcgccgcga	ggggagccgg	23940
gacgacacgt	cttcacttgt	tgggggagct	cgccgcgcac	cgcttccgcg	ctcgggggtg	24000
gtttctgcgt	gtctctcttc	cggaactggc	atttctctct	ccatagggca	gaaaaaatc	24060
atggagtcag	tcgagaagaa	ggacaagcta	accgccccct	ctgagtttgc	caccaccgcc	24120
ttccactgatg	ccgcgaacgc	gcttaccacc	ttcccgctcg	agggaccctc	gcttgaggag	24180
gaggagtgca	ttatcagaga	ggaccacagt	tttgtaacgc	aagacagcga	ggaccgctca	24240
gtaccacacg	aggtataaaa	gcaagaccag	gacacagcga	agggcaaacg	ggaacagctc	24300
ggcgcggggg	acgaagggca	tgccgactac	ctagatgtgg	gagacgcagc	cgtgttgagg	24360
cactctgcagc	ccaggtgcgc	catttctctg	gacgcgtttc	aagagggcag	cga'tgtgcc	24420
ctcgcgatag	cggatgttcg	ccttgcctac	gaacgcaccc	tattctcaac	cgctgttacc	24480
ccccaaacgc	aagaaaacgg	cacatgcgag	ccccaccgcg	gcttcaactt	ctaccccgta	24540
tttgcgtgct	cagaagtgct	tgccacctat	cacatctttt	ttccaaaactg	caagataacc	24600
ctatccttgc	gtgccaacgc	cagcgcagcg	gacaagcagc	tgcccttgcc	gcagggcgct	24660
gtcatacctg	atattcgctc	gctcaacgaa	gtgccaaaaa	ttcttgaggg	ttcttgagc	24720
gcaggaagag	ggcgcgcaaa	cgctctgcaa	caggaaaaac	gcgaaaatga	aagctactct	24780
ggagctgttg	tggaactcga	gggtgacaac	ggcgccctag	cgttactaca	acgcagcatc	24840
gaggttcaacc	actttgccta	cccgcgactt	aacctacccc	ccaaggtcat	gagcacagtc	24900

atgagtgatg	tgatcgtg	cgtgctgc	cccttgaga	gggatgcaa	tttgcaagaa	24960
aaacacagag	aggccctacc	cgcagttggc	gacgagcagc	tagccgcctg	gcttcaaacg	25020
cgccgagcctg	ccgacttgga	ggagcgacgc	aaactaatga	tggccgcagt	gctcgttacc	25080
gtggagcttg	agatctgca	gggtttcttt	gctgaccctg	agatgcagcg	caagcttagag	25140
gaacacattgc	actacacctt	tgcacagggc	tacgtacgcg	aggcctgcga	gatctccaac	25200
gtggagctct	gcaacctggg	ctcctacctt	ggaattttgc	acgaaaaccg	ccttggggcaa	25260
aaactgtcttc	attccacgct	caagggccgag	gcggcccgcg	actacgtccg	cgactcgctt	25320
tacttatcttc	tatgctacac	ctggcgagacg	gcatggggcg	tttgggcagca	gtgcttggag	25380
gagtgcaacc	tcaaggagct	gcagaaactg	ctaaagcaaa	acttgaaggga	cctatggagc	25440
gccttcaacg	agcgctccgt	ggcgcgacac	ctggcgggaca	tacttttccc	cgaacgcctg	25500
cttaaaaccc	tgcaacaggg	tctgccagac	ttcaccagtc	aaagcatggt	gcagacactt	25560
agggaactta	tctatagagc	ctcaggaatc	ttgcccgcca	cctgctgtgc	acttccctagc	25620
gactttgtg	ccattaaagta	cgcggaatgc	ctcccgccgc	tttggggccca	ctgctacctt	25680
ctgcagctag	ccaactacct	tgcctaccac	tctgacataa	tggaaagcgt	gagcggtgac	25740
gggtctactgg	agtgtcaactg	tgcctgcaac	ctatgcaccc	cgcaccgcctc	cctggtttgc	25800
aattcgcagc	tgcttaacga	aagtcaaat	atcggtacct	ttgagctgca	gggtccctctg	25860
cctgacgaaa	agtcgcggcg	tccggggttg	aaactcactc	cggggctgtg	gacgtcgctt	25920
taccttctgca	aattttgtacc	tggagactac	cacgcccacg	agattaggtt	ctacgaagac	25980
caatcccgcc	gcaccaatgc	ggagcttacc	gcttcgctca	taccacaggg	ccacatcttt	26040
ggccaattgc	aaagcactcaa	caaagcccg	caagagtttc	tgtacgaaa	ggggcggggg	26100
gtttacttgg	acccccagtc	cggcgaggag	ctcaacccaa	tcccccgcc	gcgcgacgcc	26160
tatcagcagc	agcccgggcg	ccttgccttc	caggatggca	cccaaaaaga	agctgcagct	26220
gcgcgcgcca	cccccgagc	aggaggaata	ctgggacagt	agggcagagg	aggttttggg	26280
cgaggaggag	gagagcatga	tggaaagact	ggagagccta	gacgaggaag	cttccagagc	26340
cgaagagggt	tcagacgaaa	caccgtcacc	ctcggtgcga	ttccctcgc	cgccgcacca	26400
gaaatcgcca	acgggttcca	gcattggctac	aaactccgct	cctcaggcgc	cgccgcgcat	26460
gcgcgtctgc	cgaccaaac	gtagatggga	caccactgga	accaggggcg	gtaagtccaa	26520
gcagcccgcc	ccgttagccc	aaagcaaca	acagcgccaa	ggctaccgct	catggcgcg	26580
gcacaagaac	gccatagttg	cttgcctgca	agactgtggg	ggcaacatct	ccttcgccc	26640
ccgctttctt	ctctaccatc	acggcgctgc	cttccccggt	aaatcctcgc	attactaccg	26700
tcactcttac	agccctactc	gcaccggcgg	cagcggaacg	ggcagcaaca	gcagcggcca	26760
cacagaagca	aaggcgaccg	gatagcaaga	ctctgacaaa	gcccagaaga	tccacagcgg	26820
cggcgacgag	aggagggaga	gcgctgcgtc	tggcgcccaa	cgaaccgcta	tcgacccgcg	26880
agcttagaaa	caggattttt	ccactctgtc	atgctatat	tcaacagagc	agggggccag	26940
aaacaagact	gaaataaaaa	aacaggtctc	tgcgatccct	cacccgcagc	tgcctgtatc	27000
acaaaagcga	agatcagctt	cggcgcaacg	tgggaagcgc	ggaggctctc	ttcagttaat	27060
actcgcgcct	gactcttaag	gactagtttc	gcgccttttc	tcaaatitaa	gcgcgaaaaa	27120
taagtcatct	cacggggcca	caccggcgcg	cagcacctgt	cgctcagccc	attatgagca	27180
aggaaatttc	cagccctcac	atgtggagtt	accagccaca	aatgggactt	gcggctggag	27240
ctgcccaga	ctactcaacc	gaaataaact	acatgagcgc	gggaccccac	atgatatccc	27300
gggtctctgc	taattccgcc	caccgaaacc	gaattctctt	ggaaacggcg	gctattacca	27360
ccacacctgc	gaataacctt	aatcccccta	gttggcccg	tgccctgggt	taccaggaaa	27420
gtcccgcctc	cacactgtg	gtacttccca	gagacgcca	ggccgaagtt	cagatgacta	27480
actcaggggc	gcagcttgcg	ggcggttttc	gtcacagggt	cggtgcgcc	gggcagggta	27540
taactcaact	gacatccaga	ggcgagagta	ttcagctcaa	cgacgagctc	gtgagctcct	27600
cgcttggtct	ccgtccggac	gggacatttc	agatccggcg	cgccggccgt	ccttcatcca	27660
gcgctctgca	ggcaattccta	actctgcaga	cctcgtcttc	tgaacgcgcg	cttggaggca	27720
ttggaaactct	gcaatttatt	gaggagtttg	tgcctcgtt	ctactttaac	ccctctctcg	27780
gacccctccg	ccaactccgc	gatcaattta	ttcctaactt	tgcagcgcta	aaggactctg	27840
cggacggcta	cagctgaagt	ttaagtggag	aggcagagca	actcgcctg	aacacactgg	27900
tcaccctgtc	cgcccaacag	tgtcttgccc	gcgactccg	tgaattttgc	tactttgaat	27960
tgcccgagga	tcatatcgag	ggcccgcgcg	acggcgctcg	gcttaccgcc	cagggagagc	28020
tgcgcgctag	ctctatggcg	gagtttaacc	agccgcccc	gctagttag	cgggacaggg	28080
gacctctgtg	tctcaactgt	atttgcacat	gtcctaacct	tggattacat	caagatcttt	28140
gttgccactc	ctgtgctgag	tataataaat	acagaaatta	aaatatactg	gggctcctat	28200

cgccactctg	taaacgccac	cgctctccac	cgcccaagca	aaccaaggcg	aaccttacct	28260
ggctactttta	acatctctcc	ctctgtgatt	tacaacagtt	tcaaccagga	cgaggatggt	28320
ctacgagaga	acctctccga	gctcagctac	tcctacagaa	aaaacaccac	ctctcttacc	28380
tgcggggaac	gtacgagctg	gtccacggcc	gctgcaccac	acctaccgcc	tgcacgttaa	28440
ccagactttt	tcgggacaga	cctcaataac	ctgttttacc	agaaacggag	ggcggcttag	28500
aaaaaccctta	gggtatttag	ccaaaggcgc	agctactgtg	gggttttatga	acaattccaag	28560
caactctacg	ggctatttcta	attcagggtt	ctctagaattc	gggggtgggg	ttattctctg	28620
tcttgtgatt	ctctttattc	ttatactaac	gcttctctgc	ctaaggctcg	cgccgtgctg	28680
tgtgcacatt	tgcattttatt	gtcagctttt	taaacgcgtg	ggcgccacc	caagatgatt	28740
aggtacataa	tcctagtttt	actcaccctt	gcgtcagccc	acggtaccac	ccaaaagggt	28800
gatttttaagg	agccagcctg	taattgttaca	ttcgacgtg	aagctaata	gtgcaccact	28860
cttataaaat	gcacacaga	acatgaaagg	ctgcttattc	gccacaaaa	caaaattggc	28920
aagtatgctg	ttatgtctat	ttggcagcca	ggtagacata	cagagtataa	tgctacagt	28980
ttccagggtta	aaagtcatata	aacttttatg	tatacttttc	catttttatga	aatgtgcgac	29040
attaccatgt	acatgagcaa	acagtataag	ttgtggcccc	cacaaaattg	tgtggaaaac	29100
actggcactt	tctgtgcacg	tgctattgta	attacagctg	tcgctttggg	ctgtacccta	29160
ctctatatta	aatacaaaa	cagacgcagc	tttattgagg	aaaagaaaat	gccttaattt	29220
caaatagttt	aacagtaatt	tcaccactaa	ctgctttact	cgctgtctgc	aaaaacaatt	29280
caaaaagtta	gcattataat	tagaattagga	tttaaacccc	ccggtcattt	ctcgtcaat	29340
accattccccc	tgaaacaatt	actctattgt	ggatattgctc	cagcgctaca	accttgaagt	29400
caggctctctg	ctgtctgacg	atctgacttt	ggccagcacc	tgctcccgcg	atttgttcca	29460
gtccaactac	acgacccac	cctaacagag	atgaccaaca	caaccaacgc	ggccgcgcgt	29520
accggactta	catctaccac	aaatacacc	caagttttct	cccttgtcaa	taacttggat	29580
taacttgggca	tgtgtgtggt	ctccatagcg	cttatgtttg	tatgccttat	tattatgttg	29640
ctcatctgct	gcctaaggcg	caaacgcgcg	cgaccaccca	tcctatagtc	ctcatgtgtg	29700
ctcacaccac	acaatgatgg	aatccataga	ttggacggac	tgaacacata	gtctttttct	29760
cttcacgtat	gattaaatga	gacatgattc	ctcgagtttt	tatatattct	acccttgggt	29820
cgcttttttg	tgcgtgctcc	acattgtgctg	cggttttctc	catcgaagta	gactgcattc	29880
cagcctctac	agtctatttg	ctttacggat	ttgtcaacct	cacgctcaat	tgcagcctca	29940
tcactgtgtg	catcgctctt	atccagtgca	ttgactgggt	ctgtgtgcgc	tttgcatatc	30000
tcagacacca	tcgccagtac	agggacagga	ctatagctga	gcttcttaga	attctttaa	30060
tatgaatttt	actgtgactt	ttctgtctat	tatttgacc	ctatctgcgt	tttgttcccc	30120
gacctccaag	cctcaaaagc	atatatcatg	cagattoact	cgtatatgga	atattccaag	30180
ttgtcacaat	gaaaaaaagc	attctttccga	agcctgggta	tatgcaata	tcctgtgtat	30240
gggtgttctgc	agtaacctat	tagccctagc	tatatatccc	taccttgaca	ttgtcctggaa	30300
acgaatagat	gccatgaact	acccaacttt	ccccgcgcc	gctatgcttc	caactgcaaca	30360
agttgttgcc	ggcggctttg	tccagcccaa	tcagcctgc	cccactctc	ccaccccac	30420
tgaattcagc	tacttttaac	taacaggagg	agatgactga	cacctagat	ctagaataag	30480
acggaattat	tacagagcag	cgccctctga	aaagacgcag	ggcagcgccc	gagcaacagc	30540
gcgtgaatca	agagctccaa	gacatggtta	acttgcaaca	gtgcaaaaag	ggatattctt	30600
gtctggtaaa	gcctcagcaaa	gtcacctacg	acagtaatac	accggacac	cgccattgct	30660
acagttggcc	aaccagcgtg	cagaaattgg	tggctatggt	gggagaaaag	cccattacca	30720
taactcagca	ctcggtagaa	acccaaggct	gcattcaact	accttgtcaa	ggactgtagg	30780
atctctgcac	ctttataaag	acctgtggcg	gtctcaaaag	tctatttccc	tttaactaat	30840
aaaaaaaat	aataaaggat	cacttactta	aaatcagtta	gcgaatttct	gtccagttta	30900
ttcagcgagca	ctcctctgcc	ctctggattt	ctctggattt	gcagcttctc	ctcgggtgca	30960
acttttctcc	acaaattaaa	tggaattgtca	gtttctctct	gttctgtctc	atccgcaccc	31020
actattctca	tgttgttgca	gatgaagcgc	gcaagacgtg	gcaagataac	cttcaacccc	31080
gtgtatccat	atggacacgga	aacccgtctc	ccaactgtgc	cttttcttca	tcctcccttt	31140
gtatcccccac	ttgtgttttca	agagagctcc	ctcgggtgac	tccttttgcc	ctctaccgaa	31200
ctctatggtta	ctctcaatgg	catgcttgcc	ctcaaaatgg	gcaacggcct	ctctctggac	31260
gaggcgggca	acctaacttc	ccaaaatgta	accactgtga	ccgacactct	caaaaaaac	31320
aagtcaaaaca	taaacactgga	aatatctcga	ccccacacag	ttacctcaga	agcccttaact	31380
gtggctggcg	cgcaacactc	aatggtgcgc	ggcaacacac	tcaccatgca	atcacaggcg	31440
ccgctaaccg	tgcacagctc	aaaacttagc	attgcaaccc	aggagccctc	cacagtgcta	31500

gaaggaaagc	tagccttgca	aacatcaggc	ccctccacca	ccaccgatag	cagtagccctt	31560
actatacactg	ctccaccoc	tctaaactact	gcactggta	gcttggccat	tgacttgaaa	31620
agagccattc	atacaacaaa	tggaaaacta	ggactaaatg	acggggctcc	tttgcatgta	31680
acagacgacc	taaacacttt	gaccgtagca	actggtccag	gtgtgactat	taataactat	31740
tcccttgcaa	ctaaagttac	tggagccttg	ggttttgatt	cacaaggcaa	tatgcaactt	31800
aatgttagacg	gaggactaag	gattgattct	caaaacagac	gccttatact	tgatgttagt	31860
tatocgtttg	atgtccaaaa	ccaaactaaat	ctaagactag	gcagggcccc	tctttttata	31920
aactcagccc	acaacttgga	tattaaactac	aacaaggccc	ttacttgggt	tacagcttca	31980
aaacaattcca	aaaagcttga	ggttaaccta	agcacttgca	aggggttgat	gtttgacgct	32040
acagcccatag	ccattaatgc	aggagatggg	cttgaatttg	gttcaactaa	tgccaccaaac	32100
acaaatcccc	tcaaaacaaa	aattggccat	ggcctagaat	ttgattcaaa	caaggctatg	32160
gttctctaaac	taggaactgg	ccttagtttt	gacagcacag	gtgccaattac	agttagaacc	32220
aaaaataatg	ataagctaac	tttgtggacc	acaccagctc	catctcctaa	ctgttagacta	32280
aatgcagaga	aagatgctaa	actcaccttg	gtcttaacaa	aatgtggcag	tcaaatcactt	32340
gctacagttt	cagttttggc	tgttaaggcc	agtttggctc	caatatctgg	aacaggtcaa	32400
aggtgctcatc	ttatataaag	atttgcagaa	aatggagctc	tactaaacaa	ttccttccgt	32460
gaccagaagt	atttggaaact	tagaaatgga	gatcttactg	aggcacagc	ctatacaaac	32520
gcgttggtag	ttatgctctaa	cctatcagct	tatccaaaat	ctccggttaa	aactgccaaa	32580
agtaacactg	tcagtcacagt	ttacttaaac	ggagacaaaa	ctaaacctgt	aacacatacc	32640
attacactaa	acgtgtcacac	ggaaacagga	gacacaaact	caagtgcata	ctctatgtcta	32700
ttttctgggg	actcgttctgg	ccacaactac	attaatgaaa	tattttgccac	actcctttac	32760
actcttttcat	acatgtcccca	agaataaaga	atcgttttgg	ttatgtttcca	acgtgtttat	32820
ttttcaattg	cagaaaattt	caagtcattt	ttcattccag	agtatagccc	accaccacca	32880
tagctttatac	agatcacogt	accttaatac	aaactcacaga	acctcatgat	tcaactctgc	32940
acctctctccc	caacacacag	agttacacag	cccttctccc	cgccgtggct	taaaaagcat	33000
catatcatgg	gtaacagaca	tattcttagg	tgttatattc	cacacggttt	ccgtctcgagc	33060
caaacgctca	tcagtgatat	taataaaact	cccgggcagc	tcacttaagt	tcagtctgct	33120
ctccagctgc	tgagccacag	gctgctgtcc	aaacttgcgt	tgtttaacgg	ggcggaagg	33180
agaagtccac	gctacatagg	gggtagagtc	ataactcgtc	atcaggatag	ggcgttggtg	33240
ctgcagcagc	gcgcgaataa	actgctgcgc	ccgcgcctcc	gtcctgcagc	aatacaacat	33300
ggcagtggtc	tctctcagca	tgattcgcac	cgcccgacgc	ataaggcgcc	tgttctctcg	33360
ggcacagcag	cgcaacctga	tctcacttaa	atcagcacag	taactgcagc	acagcaccac	33420
aatattgttc	aaaaatccac	agtgcaagcc	gctgtatcca	aaagtcatgg	cggggaccac	33480
agaacccacg	tggccatcat	accacaagcg	caggtagatt	aaagtggcag	ccctcataaa	33540
cacgtcggac	ataaacatta	ctctttttgg	catgttgtaa	ttcaccacct	cccggtacca	33600
tataaacctc	tgattaaaca	tggcgccatc	caccacatcc	ctaaaccagc	tggccaaaaa	33660
ctgcgcgcgc	gctatacact	gcagggaacc	gggactggaa	caatgcagct	ggagagccca	33720
ggactcgttaa	ccatggatata	tcatgctcgt	catgatatac	atgttggcac	aaacagcgca	33780
cacgtgcata	cacttctcca	ggattacaag	ctctctccgc	gttagaacca	tatcccgagg	33840
aaacaaccat	tcttgtaatca	gcgtaaatcc	acacactgac	ggaagacctc	gcacgttaact	33900
cacgttgctgc	atttgtcaag	tgttaacatt	gggcagcagc	ggatgatcct	ccagttaggt	33960
agcgcgggtt	tctgtctcaaa	aaggaggtag	acgatcccta	ctgtacggag	tgcgccaga	34020
caaccagagat	cgtgttggtc	gtagtgtcat	gccaaatgga	acgcgggacg	tgatcataat	34080
tccctgaagca	aaacacaggt	cgggcgtgac	aaacagatct	ggctctccgc	tctcgcgcgt	34140
tagatcgtcc	tgttagtagt	tgttagtata	tccactctct	caaaagcatc	agggccccc	34200
tggcttcggg	tctcatgtaa	actccttcac	gcgcgcgtgc	ctgtataaca	tccaccacgc	34260
cagtaataagc	caacacccag	caactcacac	attcgttgct	cgagtcaaac	acggggaggag	34320
aaagacagaga	tgaagcaaac	atgttttttt	ttttattcca	aaagattatc	caaaaactca	34380
ctcacgtccca	cttattaagtg	aacgcgctcc	ctccggttgc	cgtgtgtcaa	ctctacagcc	34440
aaagacagaga	taatggcattt	tgttaagatg	gtccacaatg	ctcccaaaag	gcacaaagcgc	34500
ctcacgtccca	agtgtagccta	aaggctaaac	ctctcagggg	gaaatctctc	tataaacatt	34560
ccagcacctt	caacctcacc	caataaatc	tcatctcgcc	acctctctca	tatactctca	34620
agcaaatccc	gaatattaag	tccggccatt	gtaaaaatct	gtctccagagc	gccctccacc	34680
ttcagcctca	agcagcgaaat	catgattgca	aaaaatccag	ttctccacag	acctgtataa	34740
gattcaaaag	cggaaacatta	acaaaataac	cgcgactccc	taggttccctt	cgcaggggca	34800

gctgaacata	atcgtgcagg	tctgcacgga	ccagcgccgg	cacttccccc	ccagggaacct	34860
cccaaaaaga	acccacactg	attatgacac	gcatactcgg	agctatgcta	accagcgtag	34920
tgcagagtgta	agctttgtgt	catggggcgc	gatataaaat	gcaaggtgct	gctcaaaaaa	34980
tcaggcaaaag	cctcgccgaa	aaaagaaagc	acatcgtagt	catgctcatg	cagataaaag	35040
caggtaagct	cgggaacac	cacagaaaaa	gacacattt	ttctctcaaa	catgtctcgc	35100
ggtttctgcga	taaaacacaa	ataaaataac	aaaaaaacat	ttaaacatta	gaagcctgtc	35160
ttacacacagg	aaaaaacac	cttataagca	taagacggag	tcaggccatg	cggcggtgac	35220
cgttaaaaaa	ctggtcacgc	tgattaaaaa	gcacacacga	cagctcctcg	gtcatgtcgc	35280
gagtcataat	gtaagactcg	gtaaacacat	cagggtgatt	catcggtcag	tgctaaaaag	35340
gcaccgaaat	agcccgccgg	aatacatacc	cgcaggcgta	gagacaacat	tacagccccc	35400
ataggaggta	taacaaaatt	aataggagag	aaaaacacat	aaacacctga	aaaacccctc	35460
tgccataggca	aaatagcacc	ctcccgctcc	agaacacat	acagcgcttc	acagcgccag	35520
cctaacagtc	agccttacca	gtaaaaaaga	aaacctatta	aaaaaacacc	actcgccacg	35580
gcaccagctc	aatcagtcac	agtgtaaaaa	agggccaaat	gcagagcgag	tatatatagg	35640
actaaaaaat	gacgttaacg	ttaaagtcca	caaaaaacac	ccagaaaaacc	gcacgcgaac	35700
ctacgccagc	aaacgaaagc	caaaaaaccc	acaacttctc	caaatcgcta	cttcggtttt	35760
cccacgctac	gtaacttccc	attttaagaa	aactacaatt	cccaacacat	acaagttact	35820
ccgcctctaa	acctagctca	ccgcgccctg	tcccacgccc	cgcgccacgt	cacaactctc	35880
acccctctat	tatcatattg	gcttcaatcc	aaaataaggt	atatatttga	tgatg	35935

<210> 9

<211> 35935

<212> DNA

<213> Adenovirus serotype 5

<400> 9

catcataaat	aatatacctt	attttggatt	gaagccaata	tgataatgag	ggggtggagt	60
ttgtgacgtg	gcgcggggcg	tgggaacggg	gcgggtgacg	tagtagtgtg	gcgggaagtgt	120
gatgtgtgcaa	gtgtggcggg	acacatgtaa	gcgcaggatg	tgccaaaagt	gacgtttttg	180
gtgtgcggcg	gtgtacacag	gaagtgtcaa	ttttcgcgcg	gttttaggcg	gatgtttagt	240
taaatttggg	cgtaaccgag	taagatttgg	ccattttcgc	gggaaaactg	aataagaggga	300
agtgaaatct	gaataatttt	gtgttactca	tagcgcgtaa	tatttgtcta	gggccgcggg	360
gactttgacc	gtttacgtgg	agactcgccc	aggtgttttt	ctcaggtgtt	ttccgcgttc	420
cggttcaaa	ttggcgcttt	attattatag	tcagctgacg	tgtagtgtat	ttataccggc	480
tgagtctctc	aagagggcac	ctttgagtgc	cagcgagttag	agttttctcc	tcgcgacggc	540
tcgcacacgc	ggactgtaaa	tgagacatat	tatctgccac	ggaggtgtta	ttaccgaaga	600
aatggcgccc	agtccttttg	accagctgat	cgaagaggta	ctggctgata	attcttccac	660
tcctagacct	tttgaaccac	ctacccttca	cgaactgtat	gatttagacg	tgacggcccc	720
cgaagatccc	aacgagggag	cggtttcgca	gatttttccc	gactctgtaa	gtttggcgggt	780
cgaggaaggg	attgaactac	ctacttttcc	gccggcgccc	ggttctccgg	agccgcctca	840
cttttccggc	cagcccgagc	agccggagca	gagagccttg	ggtccgggtt	ctatgccaaa	900
ctttgtaccg	gaggtgtatg	attctacctg	ccacgagcct	ggctttccac	ccagtgacga	960
cgaggaatga	gaggggtgag	agtttgtgtt	agattatgtg	gagcaccccg	ggcagcggttg	1020
caggtcttgt	cattatcacc	ggaggaatac	gggggaccca	gatattatgt	gttcgctttg	1080
ctatatggag	acctgtgcga	tgtttgtcta	cagtaagtga	caattatggg	cagtggggtga	1140
tagaagtgtg	gggtttgtgt	ggtaattttt	tttttaattt	ttacagtttt	gtggttctaa	1200
gaattttgtg	ttgtgatatt	tttaaaaggt	ctgtgtctgc	aaactgagcc	tgagcccgag	1260
ccagaaacgg	agcctccaag	acctaccgcg	cgctccataa	ggccgcctgc	tatcctcgaga	1320
gcgccacat	caactgtgtc	tagagaaatg	aatagtatga	cggatagcat	tgactccoggt	1380
ctttctaaaca	caectctctga	gatacacccg	gtgttcccg	ttgtccccat	taaacocagtt	1440
gcgctgagag	ttgttggggg	tgcgcaggct	gtggaaatga	tcgaggacct	gcttaacgag	1500
ctctgggcaac	cttttgcaact	gagctgtaaa	gcgcccgccc	cataaggtgt	aaacctgtga	1560
ttgcgtgtgt	ggttaaacgcc	tttgtttgtc	gaatgagttg	atgcaagttt	aataaaaggt	1620
gagataatgt	ttaacttgcg	tgccgtgtta	aatggggcgg	ggcttaaaag	gtatataatg	1680
cgccgtgggc	taactcttgt	tacatctgac	ctcatggagg	cttgggaggt	tttgggaagt	1740

ttttctgctg	tgcgtactt	gctggaacag	agctctaaca	gtacctcttg	gttttggagg	1800
ttttctgagg	gctcatccca	ggcaaaagta	gtctgcagaa	ttaaggagga	ttacaagtgg	1860
gaatttgaag	agcttttga	atcctgtggt	gagctgtttg	attctttgaa	tctgggtcac	1920
gagcgctgtg	tccaagagaa	ggtcatcaag	actttggatt	tttccaccc	ggggcgctgt	1980
gcggctgctg	ttgctttttt	gagttttata	aaggataaat	ggagcgaaga	aacctatctg	2040
aggggggggt	acctgctgga	ttttctggcc	atgcactgtg	ggagagcggg	tgtgagacac	2100
aagaatcgcc	tgctactggt	gtcttccgtc	cgcccgaggc	taataccgac	ggagagacag	2160
cagcagcagc	aggagggaag	caggcgggcg	cgccaggagg	agagcccatg	gaacccgaga	2220
gcggcgctgg	accctcgggg	atgaatgttg	tacaggtggc	tgaactgtat	ccagaactga	2280
gagcactttt	gacaaataca	gaggtatggc	aggggttaaa	gggggttaag	agggagcggg	2340
gggcttgtag	gtctacagag	gaggcttaga	atctagcttt	tagcttaatg	accgacaccc	2400
tctctgagtg	tattactttt	caacagatca	aggataattg	cgctaatgag	cttgatctgc	2460
tggcgagaaa	gtattccata	gagcagctga	ccacttactg	gctgcagcca	gggagtgatt	2520
ttaggaggcg	tattagggta	tatgcaaaag	tggcacttag	gccagattgc	aagtacaaga	2580
tcagcaaaact	tgtaaatatc	aggaatttgt	gctacatttc	tgggaacggg	gcggaggttg	2640
agatagatag	ggaggatagg	gtggccttta	gatgtagcat	gataaatatg	tggcgggggg	2700
tgctttggcat	ggacgggggt	gttattatga	atgtaagttt	tactggcccc	aatttttagc	2760
tacggctttt	ctctggccaat	accaacctta	tcttacacgg	tgtaaagtct	tatgggttta	2820
acaataacctg	tgtggaagcc	tggaccgatg	taagggttgc	gggctgtgce	ttttactgct	2880
gctgggaagg	gggtgtgtgt	cgccccaaaa	gcaggggtct	aattaaagaa	tgccctcttg	2940
aaaggtgtac	ctctgtgtatc	ctgtctgagg	gtaactccag	gggtgcgccac	aagtgtgctc	3000
cgactgtgtg	ttgcttcatg	ctagtgaaaa	cgctgtgctg	gattaagcat	aacatggtat	3060
gtggcaactg	cgaggacagg	gccttccaga	tgtctgacct	ctcggacggc	actgtccacc	3120
tgtctgaagac	catctacgta	gccagccact	ctcgcaaggc	ctggccagtg	tttgagcata	3180
acatactgac	ccgctgttcc	ttgcatttgg	gtaacaggag	gggggtgttc	ctaccttaac	3240
aatgcgaatt	gagtcacact	aagatatctc	ttgagcccca	gagcatgtcc	aaggtgaacc	3300
tgaacggggg	gtttgacatg	accatgaaga	tctgggaagg	gctgaggtac	gatgagacc	3360
gcaccagggt	cagaccctgc	gagtggtggc	gtaaacatat	taggaaccag	ctgtgatgac	3420
tggatgtgac	cgaggagctg	aggcccgatc	acttggtgct	ggcctgcacc	cgctctgagt	3480
tgtgctctag	cgatgaagat	acagatttag	gtactgaaat	gtgtggcgct	ggcttaagggt	3540
tgggaagaaa	tataaaggt	gggggtctta	tgtagttttg	tatctgtttt	gcagcagccg	3600
cgccgcctac	gagcacaac	tctgtttgat	gaagcatttg	gagctcatat	ttgcaaacgc	3660
gcattgcctcc	atgggcgggg	gtgcgtcaaga	atgtgatggg	ctccagcatt	gatgtgtccc	3720
cgctcctcgcc	cgcaaaactct	actaccttga	ctacagagac	cggtgtctga	acgcgcttgg	3780
agactgcagc	ctcccgccgc	gcttccagcg	ctgcagccac	cgccccggg	atgtgtactg	3840
acttttgtct	cttgagcccg	cttgcaagca	gtgcagcttc	ccgttcatcc	gcccgcgagt	3900
acaagttgac	ggctcttttg	gcacaattgg	attctttgac	cggggaactt	aagtgtgttt	3960
ctcagcagct	gttggagtctg	cgccagcagg	ttctgtccct	gaaggcttcc	tcccttccca	4020
atgcgggtta	aaaacataaat	aaaaaacca	actctgtttg	gatttggatc	aagcaagtgt	4080
cttgtctgct	ttatttaggg	gttttgcgcg	cgccgttaggc	ccgggaccag	cggtctcggt	4140
cgttgagagc	ctgtgttatt	ttttccagga	cggtgtaaa	gtgactctg	atgttccagt	4200
acatgggcat	aagcccgctg	ctggggtgga	ggtagcacca	ctgcagagct	tcatgtctgc	4260
gggttggtgt	ttagatgac	cagtcgtagc	aggagcgctg	ggcgtggtgc	ctaaaaaagt	4320
ctttcagtag	caagctgatt	gccaggggca	ggcccttggt	gtaagtgttt	acaaagcggt	4380
taagctgaga	tgggtgcata	cgltggggta	tgagatgcat	cttggaactg	atttttaggt	4440
tggctatggt	ccagcccata	tccctccggg	gattactgtt	gtgcagaacc	accagcacag	4500
tgtatccggt	gcacttgagg	aatttgtcat	gtagcttaga	aggaatctgc	tgggaagaact	4560
tgggagcgcc	cttgtgacct	ccaagatttt	ccatgcattc	gtccataatg	atggcaatgg	4620
gcccaacggc	ggcgcccgctg	gggaagatat	ttctgggata	actaaogcta	tagttgtgtt	4680
ccaggagtag	atcgctcatg	gccattttta	caaagcggg	cgggaggggtg	ccagacttgc	4740
gtataatggt	tccactccgg	ccaggggcgt	agttaccctc	acagatttgc	atcttccccc	4800
cttttagtgc	agatgggggg	atcatgtcta	ctcgggggg	gatgaagaaa	acggtttccg	4860
gggttagggga	gatcagctgc	gaagaagaca	ggttctctag	cagctgcgac	ttaccgcagc	4920
cggtggggccc	gttaatacaca	ccattaccgc	gggtcgaactg	gtagttaaga	gagctgcagc	4980
tgcctctatc	ctctgacagg	ggggccactt	cgtaagcat	gtccctgact	cgcatgtttt	5040

ccctgaccaa	ataccgcaga	aggcgctcgc	cgccacgcga	tagcagttct	tgcaaggaa	5100
caaaagtctt	caacgggttg	agaccgctcg	ccgttaggcat	gcttttgagc	gttttgaccaa	5160
gcagttccag	cgcgtccacc	agctcggtca	cctgctctac	ggcatctcga	tccagcatat	5220
ctccctgttt	cgccgggttg	ggcggtcttc	gctgtaccgc	agtagtcggt	gctcgtccag	5280
acggggccagg	gtcatgtctt	tccacggggc	cagggtctctc	gtacgcgtag	tctgggtcac	5340
ggtagaaggg	tgccgtccgg	gctcgcgcgt	ggccagggtg	cgcttgaggc	tggtctctgt	5400
ggtagctgaag	cgctgcgggt	cttcgcgcct	cgctcggccc	aggtagcatt	tgaccatggg	5460
gtcatagtcc	agccctccgc	cgccgtggcc	cttgccgcgc	agcttgccct	tggaaggagg	5520
gcgcgcacgag	ggcgagtgca	gacttttgag	ggcgtagagc	ttggggcgca	gaataccgca	5580
ttccggggag	taggcctccg	cgccgcaggc	cccgacagag	gtctcgccatt	ccacgagcca	5640
gttagctctc	ggccgttcgg	ggcctcaaaac	cagggttccc	ccatgctttt	tgatgcgttt	5700
cttacctctg	gttttccatg	gcccgtgtcc	acgctcgggt	acgaaaaggc	tgctcgtgtc	5760
cccgctataca	gacttgagag	gcctgtcttc	gagcggtgtt	cccggtctct	ccctgatatag	5820
aaactcggac	cactctgaga	caaaggctcg	cgctccaggcc	agcacgaagg	aggctaaagt	5880
ggaggggtag	cggtcgttgt	ccactagggg	gtccactcgc	tccagggtgt	gaagacacat	5940
gtcgccctct	tcggcatcaa	ggaagggtgat	tggtttgtag	gtgtaggcca	cgtgacccgg	6000
tgttcctgaa	ggggggctat	aaaagggggt	ggggggcggt	tcgtctctcac	tctcttccgc	6060
atcgctgtct	cgaggggcca	gctgttgggg	tgagtactcc	ctctgaaaaa	cgggcatgac	6120
ttctgctcta	agattgtcac	tttccaaaaa	cgaggaggat	ttgatattca	ctctggccgc	6180
ggtagtgctt	ttgagggttg	ccgcatacat	ctggtcagaa	aaagacaatc	ttttgtgttc	6240
aagctgcgtg	gcaaaagacc	cgtagagggc	gttggacagc	aacttgccga	tggaagcgag	6300
gggttggttt	tgtcgcgat	cgccgcgcct	cttgccgcgc	atgttttagct	gcacgtattc	6360
cgccgcacag	cagcccatct	cggaagaagac	gggtggtgac	tcgtcggcca	ccagggtgac	6420
cgcccaaccg	cggttgtgca	gggtgacaaag	gtcaaccgctg	gtggctacct	ctccggtgag	6480
cgctcgtgtg	gtccacagag	ggcgccgcgc	cttgcgcgag	cagaatggcg	gtagggggtg	6540
tagtctcgtc	tcgtccgggg	ggctcgcgtc	cacggtaaaag	accccgggca	gcagggcgccg	6600
gtcgaagttag	tctatcttgc	atccttgcaa	gtctagcgcc	tgctgccatg	cggggcgccg	6660
aagcgccgcg	tcgtatgggt	tgagtggggg	accctatggc	atgggggtgg	tgagcgccgga	6720
ggcggtacatg	ccgcaaatgt	cgtaaacgta	gagggtctct	ctgagtattc	caagatatgt	6780
agggtagcatc	cttccaccgc	ggatgctggc	gcgcacgtaa	tcgtatagtt	cgtgcgaggg	6840
agcgagggag	tcgggaccga	ggttgctacg	ggcgggctgc	tctctcggga	agactatctg	6900
ctcgtgaagt	gcattgtgag	tgatgatata	ggttggaacg	tggaagacgt	tgaagctggc	6960
gtctgtgaga	cctaccgcgt	cacgcacgaa	ggaggcgtag	gagtcgccca	gcttgttgac	7020
cagctcggcg	gtgacctgca	cgcttagggc	gcagtagtcc	agggtttctc	tgatgatgtc	7080
atacttatacc	tgctcccttt	ttttccacag	ctcgcgggtg	aggacaaact	cttcgcggtc	7140
ttttccagta	ctttggatcg	gaaccctcgc	ggcctccgaa	cggtaagagc	ctagcatgta	7200
gaactgggtg	acggcctggg	aggcgacgca	tcctttttct	acgggtagcg	cgtatgcctg	7260
cgccggcttc	cggaagcgag	tgtgggtgag	cgcaaaaggt	tccttgacca	tgacttttag	7320
gtactggtat	ttgaagtcaag	tgctgtcgca	tcgcgcctgc	tccacagaca	aaaagtcogt	7380
cgctgttttg	gaacgcggat	ttggcagggc	gaaggtgaca	tcgttgaaga	gtatctttcc	7440
cgccgcaggc	ataaagttgc	gtgtgatcgc	gaagggtccc	ggcacctcgg	aacggtttgt	7500
aattactcgg	cgccgcagca	cgatctcgtc	aaagccggtg	atgtttgtag	ccacaatgta	7560
aagttccaaag	aaagcgggga	tgcccttgat	ggaaaggcaat	tttttaagt	ctcgttaggt	7620
gagctcttca	ggggagctga	gcccgtgtct	tgaaggggcc	cagttctgaa	gatgaggggt	7680
ggagcgcgag	aatgagctcc	acaggtctacg	ggccattagc	atgtgcaggt	ggctcgcaaa	7740
ggctctaaac	ttggcgacct	tgccattttt	ttctgggggt	atgcagtaga	aggtcaagcg	7800
gtctttgtcc	cagcgctccc	atccaaaggt	cgccggttag	tcctcgccgg	cagctactag	7860
agggctcatct	ccgcggaaact	tcatgaccag	catgaaggcc	acgagctgct	tcccaaaagg	7920
ccccatccaa	gtataggctc	ctacatcgta	ggtgacaaag	agacgctcgg	tgccgagggat	7980
cgagcccgatc	gggaagaact	ggatctcccg	ccacccaatg	gaggagtggt	tattgatgtg	8040
gtgaagtagt	aagtcctctg	gacggggcga	acactcgtgc	tgccctttgt	aaaaacgtgc	8100
gcagtagctg	cagcggtgca	cgggctgtac	atcctgcacg	aggttgacct	gacgaccggc	8160
cacaaggaag	cagagtgagg	atttgagccc	ctcgctcggc	gggtttggct	gggtggtctc	8220
tacttcgctc	gctgttccct	gaaccgtctg	ctgctcggag	ggaggttacg	tggaatcgac	8280
caccacgcgc	cggcagccca	aagtcagagt	gtccgcgcgc	ggcggtcgga	gcttgatgac	8340

aaatcgccgc	agatgggggc	tgtccatggt	ctggagctcc	cgccggcgca	ggctcaggcg	8400
gagctcctgc	aggttttaact	cgcatagacg	ggtcagggcg	cgggctagat	ccaggctgata	8460
ctcaattttcc	aggggctgggt	tgggtggcgcc	gtcgatggct	tgcaagagcg	cgcattcccc	8520
cgccggcgact	acggtaacgc	gcggcgggcg	gtggcgccgc	gggggtgctc	tgagatgatgc	8580
atctaaagcg	ggtgacgcgg	gcgagccccc	ggaggttaggg	ggggctccgg	accgcgcggg	8640
agagggcgca	ggggcagctc	ggcgccgcgc	gcgggcaggga	gctggttgctg	cgccgctaggg	8700
ttgctgggca	acgcgcagac	gcggcggttg	atctcctgaa	tctggccgct	ctcgcgtgaag	8760
acgacgggcg	cgggtgagctt	gagcctgaaa	gagagttcga	cagaataaat	ttcgggtgtcg	8820
ttgacggcg	cttggcgcaa	aatctctctc	acgtctcctg	agttgtcttg	ataggcgatc	8880
tcggccatga	actgtcctat	ctcttctctc	tggagatctc	cgcgcccgcc	tcgctccagg	8940
gtggcgggcg	ggctcgttga	aatgcgggccc	atgagctgcg	agaaggcggt	gaggccctccc	9000
ctcgttccaga	cgccgctgta	gaccacggccc	ccttcggcat	cgccggcgcg	catgacaacc	9060
tgccgcagat	tgagctccac	gtgcgcggcg	aagacggcg	agtttccgag	gcgctgaaag	9120
aggttagttga	gggtggttggc	gggtgtgtct	gccacgaaga	agtacataac	ccagcgtcgc	9180
aaactgtgatt	cgttgatact	cccccaaggcc	tcaaggcgct	ccatggcctc	gtagaagtcc	9240
acggcgaaagt	tgaaaaactg	ggagtttgccg	gcgcgacacg	ttaactctct	ctccagaaga	9300
cggtatgagct	cgccgcacgt	gtcgcgcacc	tgcgcctcaa	aggetacagg	ggccttctct	9360
ctctcttcaa	tctctctctc	cataaggggcc	tcccttctct	cttctcttgg	cgccggtagg	9420
ggagggggga	cacggcgccg	acgcgcggcg	accgggagcg	ggctgcacaa	gcgctcgatc	9480
atctccccgc	ggcgacggcg	catggtctcg	gtgacggcgc	ggccgttctc	gcggggggcg	9540
agttggaaga	cgccgcccg	catgtcccg	ttatgggttg	cgccggggct	gcgatgcggc	9600
agggatacgg	cgttaacgat	gcattctaac	aatgtgtgtg	taggtactcc	gcgcgcggag	9660
gactctgagc	agttcgcctc	gaccggatcg	gaaaacctct	cgagaaggcg	gtcttaaccag	9720
tcacagctcg	aagtgaggct	gagcacccgtg	gcggggcgca	gcggggcgcg	gtcgggggttg	9780
tttcttgcgg	aggtgctgct	gatgatgtaa	ttaaagtagg	cggtcttgag	acggcggtatg	9840
gtgcacagaa	gcacacatgc	cttgggttccg	gcctgtgtaa	tgcgcaggcg	gtcggccatg	9900
ccccaggctt	cgtttttgaca	tcggcgccagg	tctttgtagt	agtccttgcat	gagcctttct	9960
acccggcactt	cttcttctct	tctctcttgt	cctgcatctc	ttgcatctat	cgctgcggcg	10020
gcggcgaggt	ttggccgtag	gtggcgccct	cttctccca	tcgctgtgac	cccgaagccc	10080
ctcatcggtt	gaagcagggc	taggttcggcg	acacgcgct	cggtaatat	ggcgtgctgc	10140
acctgcgtga	gggtagactg	gaagtcatcc	atgtccacaa	agcggttgga	tgccgcccgt	10200
ttgatggtgt	aagtgacgtt	ggccataaag	gaccagttaa	cggtcttggt	accggcgtgc	10260
gagagctcgg	gtacctgag	acgcgagtaa	gccttcgagt	caaatacgta	gtcgtttgcaa	10320
gtccgcacca	ggtactggta	tcccaccaaa	aagtcggcg	gcggctggcg	gtagaggggg	10380
cagcgtaggg	tgcccggggc	tccggggggc	agatcttcca	acataaggcg	atgatattccg	10440
tagatgtaac	tggaactcca	ggtgatgccg	gcggcggttg	tggaggcgcg	cggaagatcg	10500
cgccgcgggt	tccagatggt	gcgcagcgcc	aaaaagtcg	ccatggtcgg	gacgctctgg	10560
ccggtcaggc	gcgcgcgaatc	gttgcagctc	tagaccgtgc	aaaaggagag	cctgtaagcg	10620
gcacctcttc	cggtgtctgg	tggaataaat	cgcaagggtg	tcatggcgga	cgaccggggg	10680
tcgagccccc	tatccggcg	tcgcgcgtga	tccatcggtg	taccgcccgc	gtgtcgaaac	10740
caggtgtgcg	acgtacagaca	acggggaggt	gctccttttg	gcttctctcc	agggcgccgc	10800
gtcgtctgcg	tacgtttttt	ggccactcgc	cgccgcagcg	gtaagcggtt	aggtcgaaaa	10860
gcggaagcat	taagtggtct	gtctcctgta	gcggaggggt	tattttccaa	gggttgagtc	10920
gcgggacccc	cgggttcgagt	ctcgagccgc	ccggagctgc	gcgaacgggg	gtttgctctc	10980
cgctcatgca	agaccccgc	tgcaaaatcc	tccggaaaca	gggacagacc	ccttttttgc	11040
ttttcccgga	tgcatcccg	gttcggcgag	atggcgcccc	ctctccagca	gcggcaagag	11100
caagagcagc	ggcagacatg	cagggcaccc	tccctctctc	atggcggtc	agggggcgcg	11160
acatccggcg	ttgacgcggc	agcagatggt	gattacgaac	ccccgcggcg	ccggggcccg	11220
acatccacgg	acttgtagga	gggcgagggc	ctggcgccgc	ttagagcgcc	ctctctgtag	11280
cggtaaccaa	gggtgcagct	gaagcgtgat	acgctgaggg	cgtaactgcc	gcggcagaaac	11340
ctgtttccgc	accgcgaggg	agagggagccc	gagggatgc	gggatcgaaa	gttccacgca	11400
ggcgcgagcg	tcggcatgg	cctgaatcgc	gagcggttgc	tgccgcgagg	ggactttgag	11460
ccccagcgcc	gaaccgggat	tagtccccgc	cgccacacag	tggccggccg	cgaccttgta	11520
accgcatacg	acgcagcggt	gaaccagggg	atlaactttc	aaaaaagctt	taacaaacac	11580
gtgcgtaacg	ttgtggcgcg	cgaggagggt	gctataggac	tgaatgcatc	gtgggaacttt	11640

gtaagcgcgc	tgaggcaaaa	cccaaatagc	aagccgcctca	tggcgcagct	gttccttata	11700
gtgcagcaca	gcaggacaaa	cgaggcattc	agggatgcgc	tgctaaacat	agtagagccc	11760
gaagggcgcg	ggctgctcga	tttgataaac	atcctgcaga	gcatagtggg	gcaggagcgc	11820
agcttgagcc	tggtcgacaa	ggtggccgcg	atcaactatt	ccatgcttag	cctggggcaa	11880
ttttacgcc	gcaagatata	ccataccctt	tacgttccca	tagacaagga	ggtaaagatc	11940
gagggggttct	acatgcgcatt	ggcgcgtgaag	gtgcttacct	tgagcgacga	cctggggcgt	12000
tatcgcaacg	agcgcatcca	caaggccgtg	agcgtgagac	ggcggcgoga	gcacagcgac	12060
cgcgagctga	tgacacgcct	gcaaaaggcc	gtggctggca	cgggcagcgg	cgatagagag	12120
gcgcagctct	actttgacgc	gggcgctgac	ctgcgctggg	ccccaagccg	acgcgcctct	12180
gagggcagctg	gggcgcggac	tgggcctggcg	gtggcaccgc	cgccgcctgg	caacgtcgcc	12240
ggcgtggagg	aatatgacga	ggacgatgag	tacgagccag	aggacggcga	gtactaaagc	12300
gtgatgtttc	tgatcagatg	atgcaagacg	caacggacc	ggcgtgcg	gcggcgctgc	12360
agagcagccc	gtccggcctt	aactccacgg	acgactggcg	ccaggtcatg	gaccgcatac	12420
gtgcgtgac	tgccgcgaat	cctgacgcgt	tccggcagca	gccgcagggc	aaccggctct	12480
ccgcaattct	ggaagcgggt	gtcccgccgc	gcgcaaaacc	cacgcacgag	agagtgctgg	12540
cgatcgtaaa	cgccctggcc	gaaaacaggg	ccatccggcc	cgacgagccc	ggcctggctct	12600
acgacgcgct	gcttcagcgc	gtggctcgtt	acaaacagcg	caacgtgcag	accaacctgg	12660
acccgctggt	gggggagtg	cgcgagccgc	tggcgacgag	tgagcgccgc	cagcagcagc	12720
gcaacccagc	gtccatgggt	gcactaaaag	ccttcttgag	tacacagccc	gccaaagctg	12780
cgccgggaca	ggagagctac	accaactttg	tgagcgccat	ggcgctaagt	gtgactgaga	12840
cacgcacaag	tgaggtgtac	cagctctggc	cagactattt	tttccagacc	actgagcaag	12900
cgctgcacac	cgtaaacctg	agccagcgtt	tcaaaaactt	gcaggggcgt	tggggggtgc	12960
gggctccacg	gcgagcacgc	gcgacogtgc	ctagctttgt	gacgcccac	ctggcgctgt	13020
tgctgtgctc	aatagcgccc	ttcacggaca	gtggcagcgt	gtcccgggac	acatcaactg	13080
gtcactctgc	gcactgtgac	cgcgaggcca	taggtcagcg	gcaltggag	gacatcactt	13140
ttccagagat	tacaagtgtc	agcccgccgc	tggggcagga	ggacacgggc	agcctggagg	13200
caaacctaaa	ctacctgctg	accaaccggc	ggcagaagat	ccctctgtg	cacagttaa	13260
acacgcagga	ggagcgcatt	ttgcgctacg	tgcagcagag	cgtgagcctt	aaactgatgc	13320
gcgacggggg	aaagccccag	gtggcgcctg	acatgaccgc	gcgcaacatg	gaaccggcca	13380
tgatgtccct	aaacccggcg	tttatcaacc	gcctaatgga	ctacttgcat	cgccggcgcc	13440
ccgtgaacc	cgagtatttc	accaatgcca	tcttgaaacc	gcactgggta	cgcgccctgt	13500
gtttctacac	cgggggattc	gaggtgcocg	agggtaaagc	tggattcttc	tggagcagca	13560
tagacacagc	cgtgttttcc	ccgcaacgcg	agaacctgct	agagttgcaa	cagccgcagc	13620
aggcagaggg	ggcgttcgga	aaggaaagct	tccgcagggc	aagcagctg	tccagctag	13680
gcgctcgccg	cccgcggtca	gatgtagta	gcccatcttc	agagcttgata	gggtctctta	13740
ccagcactcg	caccaccocg	cgcgcctgc	tgggcgagga	ggagtacct	aaacactcgc	13800
tgctgcagcc	gcagcgcgaa	aaaaacctgc	ctccgcat	tcccaacac	gggatagaga	13860
gccattggga	caagtagagt	agatggaaag	gcatcgcga	ggagcacagg	gactgtccag	13920
gcccgccgcc	gccaccgcgt	cgtcaaaagg	acgaccgtca	gcgggggtctg	gtgtggggag	13980
acgattgact	ggcagacgac	agcagcgtcc	tggatttggg	aggagatggc	aaccggtttg	14040
cgacacttgc	ccccaggcgt	gggagaaagt	tttaaaaaaa	aaaaagacatg	atgcataata	14100
aaaaactcac	caaaggccatg	gcaccgagcg	tgggtttctt	tgatttcccc	ttgattagcg	14160
gcgcgcggcg	atgtatgagg	aaggtctctc	tccctctcac	gagagtgtgg	tgagccggcg	14220
gccagtggcg	ggcgcgctgg	gttctccctt	cgatgctccc	ctgggacccc	cgtttgtgac	14280
ttccggtgac	ctfcgcgcta	ccggggggag	aaacagcacc	cgttactctg	agttggcacc	14340
ccattctgac	accaccgctg	tgtacctggt	ggacaacagc	ttcaacggatg	ttggcatcct	14400
gaactaccag	aaagaccaca	gcaactttct	gaccacggtc	attcaaaaca	atgactacag	14460
cccgggggag	gcaagcacac	agacocatca	tcttgacacg	cggtcgcatc	ggggcgcgga	14520
cctgaaiaacc	atcctgcaata	ccaacatgoc	aaatgtgaac	gagttcatgt	ttaccaataa	14580
gttttaaggcg	gggtgtatgg	tgctcgcgctt	cgctactagc	gacaactcag	tggagctgaa	14640
atcacgagtg	gtggagttca	cgctgcccca	gggcaactac	tcogagacca	tgacatagaa	14700
ccttatgaac	aacgcgatcg	tggagcacta	cttgaagtgc	ggcagacaga	acggggttct	14760
ggaaagcgac	atccgggtaa	agtttgacac	ccgcaacttc	agactggggg	ttgacccctg	14820
cactgtctct	gtcatgccgt	gggtatatac	aaacgaagcc	ttccatcagc	acatcatttt	14880
gctgccagga	tgccgggtgg	acttccacca	cagccgcctg	agcaacttgt	tgggcatccg	14940

caagcggcga	cccttccagg	aggcctttag	gatacactac	gatgatctgg	agggtggtaa	15000
cattcccgca	ctgtttggatg	tggacgccta	ccaggcgagc	ttgaaagatg	acaccgaaca	15060
gggcgggggt	ggcgcaaggc	gcagcaaacg	cagtggcagc	ggcgcggaag	agaactccaa	15120
cgccggcagc	ggcgcaatgc	agccggtgga	ggacatgaac	gatcatggca	ttcgcgggca	15180
ccactttggc	acacgggggt	aggagaagcg	cgctgagggc	gaagcagcgc	ccgaagctgc	15240
cgcccccgct	ggcgcaacccg	aggctcgagaa	gcttcagaa	aaacgggtga	tcaaacacct	15300
gacagagga	acgaagaaac	cgagttaaca	cctaataaag	aatgacagca	ccttcaccca	15360
gtaccgcagc	tggtaactctg	catacaacta	cgccgaacct	cagacgggaa	tcgcgtctatg	15420
gacctcgctt	tgcactccctg	acgttaacctg	cggctcggag	cagggtcact	ggctgttgcc	15480
agacatgatg	caagaccccg	tgaaccttcg	ctccacggcg	cagatcagca	acttcocggt	15540
ggctggggcc	gagctgtgtgc	cogtgcactc	caagagcttc	tacaacgacc	aggcgcgtcta	15600
ctcccaactc	atccgcagct	ttacctctct	gacccacgtg	ttcaatcgct	ttcccgagaa	15660
ccagattttg	ggcgcccgcc	cagcccccac	catcacacc	gtcagtgaaa	acgttccctg	15720
ttctacagat	caagggaagc	taccgctgcg	caacagcatc	ggagggagtcc	aggcgagtac	15780
cattactgac	gcagagcgcc	gcacctgccc	ctacgtttac	aaggccctgg	catgatgtctc	15840
gcgcgcgctc	ctatcgagcc	gcactttttg	agcaagcatg	ttcatctcta	tatcgccagc	15900
caatactgca	ggctggggcc	tgcgcttccc	aagcaagatg	tttggcgggg	ccaaggaagc	15960
ctccgaccaa	caccagtgct	gcgtgcgccg	gcactaccgc	gcgcctctgg	ggcgccacaa	16020
acgcgcgcgc	actgggcgcga	ccacgcgtga	tgcagcccat	gcgcgctggg	tggagggaggc	16080
gcgcgaactac	agcccaacgc	cgccaccagt	gtccacagtg	gacgcggcca	ttcagacagt	16140
ggctgcggga	gcgcggcgct	atgtctaaat	gaagagacgg	cgagggcgcg	tagcagcgctg	16200
ccaccgcgcg	gcaccgcggca	ctgcgcgcga	acgcgcggcg	gcggccctgc	ttaacccgcg	16260
acgtgcaccc	ggcgacggcg	cgcccatgcg	ggccgctcga	aggctggccg	cggttatgtg	16320
cactgtcccc	ccacaggtcca	ggcgacagcg	ggccgcgcga	gcagcccgcg	ccattagtgc	16380
tatgtactca	ggctgcaggg	gcaacgtgta	ttgggtgcg	gactcggtta	cgccgctgag	16440
cgctgcgcgc	gcaccccgcc	ccccgcgcaa	ctagatgcca	agaaaaaact	acttagactc	16500
gtactgttgt	atgtatccag	cgccgcgcgc	gcgcgaacgaa	gctatgtcca	agcgcgaact	16560
caaaagaagag	atgtccaggg	tcctgcgcgc	ggagatctat	ggcccccgca	agaaggaaga	16620
gcaggattac	aagccccgaa	agctaaaagc	ggctcaaaaag	aaaaagaagag	atgatgatga	16680
tgaacttgca	gacgaggtgg	aactgcgtga	cgctaccgcg	ccagggcgac	gggtacagtg	16740
gaaaggtcga	cgctgaaaac	gtgttttgcg	acccggcacc	accgtagtct	ttacgcgcgc	16800
tgcgcgctcc	ccccgaacct	acaagcgctg	gtatgatgag	gtgtacggcg	acgaggacct	16860
ggtaggacag	gcacaagagc	gcttcgggga	gtttgcctac	ggaaaagcgc	ataaggacat	16920
gctggcgctg	cccgctggag	aggccaacct	aaacacttagc	ctaaaagccg	taaacactgca	16980
gcaggtgctg	cgccgccttg	caccgtccga	agaaaagcgc	ggcctaagcg	gcgagtctgg	17040
tgacttgcca	cccccgctgc	agctgatggt	acccaagcgc	cagcgactgg	aagatgtctt	17100
ggaaaaaatg	accgtggaac	ctgggctgga	gcccaggctc	cgcgctgcgc	caatcagcca	17160
gggtggcgccg	ggactggggc	tgcagacgct	ggagcttcag	atacccacta	ccagtagcac	17220
cagtatggcc	accgcacagc	agggcattga	gacacaaacg	ttcccggttg	ctcagcggtg	17280
ggcggtatgcc	ggcggtcgag	gggtcgctgc	ggccgcgctc	aagacctcta	cggaagtcca	17340
aacggcgccg	tggatgttcc	gcgtttcagc	ccccggcgcc	cgccgggttg	cgaggaagta	17400
cgggcgccgc	agcgcgctac	tgcgccaaata	tgcctacat	ccttccattg	cgctcaccgc	17460
cggtctatcg	ggctacaact	accccccacg	aagacagcca	actaccgcgc	gcgcgaacct	17520
caactggaaac	gcgcgcgcgc	gtcgcgctgc	ccagcccgctg	gtgcgccgga	tttcggtgcg	17580
caggttggtct	cgcgagaggc	gcagggacct	gggtgtgcca	acagcgcgct	acacccacag	17640
catcggtttaa	aagcggtctct	ttgtggtctt	tgacgatgat	gccctcaact	gcgcgctcgc	17700
tttcccggtg	ccgggattcc	gaggaagaat	gcacggtagg	agggggatgg	cgcgccacgg	17760
ctcgaggggc	ggcatgtgctc	gtgcgcacca	ccggcgggcg	ggcgctgcgc	accgtcgcat	17820
gcgcggcggtg	atccctgcccc	tccttatctc	actgatcgcc	ggcgcgatgg	gcgcgctgcc	17880
cggaattgca	tcctgtggcct	tgcaggcgca	gagacactga	ttaaaaacaa	gttgcaatgtg	17940
gaaaaatcaa	ataaaaaagt	ctggactctc	acgctcgctt	ggctcgttaa	ctatttttga	18000
gaatggaaaga	catcaacttt	gcgtctctgc	ccccgcgaca	cgcttcagcg	cgttctaatg	18060
gaaactggca	agatatcggc	accagcaata	tgagcggttg	cgccctcagc	tggggctcgc	18120
tgtggagcgg	cattaaaaaat	ttcggttcca	cgtttaagaa	ctatggcagc	aaggctcgga	18180
acagcagcac	aggccagatg	ctgagggata	agttgaaaga	gcaaaaatttc	caacaaaagg	18240

tggtagatgg	cctggccctct	ggcattagcg	gggttggtga	cctggccaac	caggcagtcg	18300
aaaataaagt	taaacagtaag	cttgatcccc	gccctccctg	agaggagcct	ccaccgggctg	18360
tgggagcagtg	gtctccagag	gggcgtggcg	aaaagcgctcc	gcgcccgac	agggagaagaa	18420
ctctgggtgac	gcaataagac	gagcctccct	cgtacgagga	ggcactaaag	caaggcctcg	18480
ccaccaccgc	tccactccgc	cccatggcta	cggagtgct	ggccagcac	acaccgctaa	18540
cgcgtgacct	gccctccccc	gccagaccgc	agcagaaacc	tgtgctgcca	ggcccgaccg	18600
ccgtttgtgt	aaccctctct	agccgcgcgt	ccctgcgcgc	gcgcccgacg	ggctccgcat	18660
cgttgccggcc	cgtagccagt	ggcaactggc	aaagcaccat	gaacagcctc	gtgggtctgg	18720
gggtgcaatc	ccagaagcgc	cgacgatgct	ctggaatagc	taacgtgtcg	tatgtgtgtc	18780
atgtatcgct	ccatgtccgc	gccagaggag	ctgctgagcc	gccgcgcgcc	cgctttccaa	18840
gatggctacc	ctctcgatga	tgccgcagtg	gtcttacatg	cacatctcgg	gccaggacgc	18900
ctcggagtagc	ctgagccccc	ggctggtgca	gtttgccccc	gccaccgaga	cgtaactcag	18960
ccctgaataac	aagtttagaa	accaccgggt	ggcgccctacg	cacgacgtga	ccacgacagg	19020
gtccaccgct	ttgacgcctg	gggttaccct	tgtggaccgt	gaggatactg	cgtaactcgt	19080
caaggcgccg	ttcaccctag	ctgtgggtga	taacgctgtg	ctggacatgg	cttccacgta	19140
ctttgacatc	cgccgcgtgc	tgacacgggg	ccctactttt	aagccctact	ctggcactcg	19200
ctacaacgcc	ctggctccca	aggggtcccc	aatcctctgc	gaatgggatg	aagctcgtac	19260
tgctcttgaa	ataaaacctag	aagaagagga	cgatgacaac	gaagacgaag	tagacgagca	19320
agctgagcag	caaaaaactc	acgtatttgg	gcaggcgccct	tattctggta	taaatattac	19380
aaaggagggt	attcaaatag	gtgtcgaagg	tcaaacacct	aaatatggcg	ataaaacatt	19440
tcacactgaa	ccctcaaatg	gagaatctca	gtgttcagaa	actgaaatta	atcctgcagc	19500
tgggagagtc	cttaaaaaag	ctaccccaat	gaacacctgt	tacggttcat	atgcanaaac	19560
cacaaatgaa	aatggagggc	aaggctattct	tgtaaagcaa	caaaaatggaa	agcttaagac	19620
tcaagtggaa	atgcaatttt	tctcaactac	tgaggcgccag	gcaggcaatg	gtgataactt	19680
gaactctaaa	gtgtgtattg	acagtgaaga	tgtagatata	gaaccccccag	acactcatat	19740
ttcttcatct	cccactatta	aggaaggtaa	ctcacagaaa	ctaattggcc	aaacattcat	19800
gcccaacagg	cctaattaca	ttgcttttag	ggacaatttt	attggtctaa	gtattataca	19860
cgacacgggt	aatatgggtg	ttctggccgg	ccaagcatcg	cagttgaatg	ctgtgttaga	19920
tttgcagagc	agaaacacag	agctttcata	ccagcttttg	cttgattcca	ttgtgttagat	19980
aaccaggtag	tttctatgt	ggaatcaggc	tgttgacagg	tatgatccag	atggtagaat	20040
tattgaaaat	catggaactg	aagatgaact	tccaattatc	tgctttccac	tgggagggtg	20100
gattataaca	gagactctta	ccaaggtaaa	acctaaaaca	ggtcaggaaa	atggatggga	20160
aaaagatgct	acagaatttt	cagataaaaa	tgaataaaga	atttggaata	attttgccat	20220
gaaatcaat	ctaaatgcca	acctgtggag	aaatttccgt	tactccaaca	tagccgtgta	20280
tttgcgcgac	ctaaataagt	acagctcttc	caacgtaaaa	atttctgata	acccaacac	20340
ctacgactac	atgaacaagc	gagtggtggc	tcgccgggta	gtggactgct	acctaaacct	20400
tggaagcagc	tggtcccttg	actatatgga	caacgtcaac	cgatttaacc	accaccgcaa	20460
tgctggccgt	cgtaaccgct	caatgttgct	gggcaatggt	cgctatgtgc	cttccacact	20520
ccagggtcct	cagaagttct	ttgccattaa	aaacctctct	ctcctgcggg	gctccataac	20580
ctacagtggt	aactctcagg	aggatgttaa	catggttctg	cagagctccc	taggaaatga	20640
ccctaagggt	gacggagcca	gcatttaagt	tgatagcatt	tgctttcgtg	ccactctctt	20700
cccatggccc	cccaacacgc	ctctccagct	tgagggccatg	cttagaagag	acaccaacga	20760
ccagctctct	aacagctatc	tctccgcgcg	caacatgtctc	taccctatac	ccgcacaacg	20820
taccaacgtg	cccatatcca	tccctccccc	caactggggc	gctttccgcg	ctgtgggctt	20880
cacgcgcctt	aagactaaag	aaaccctcat	actgggctgg	ggctacgacc	cttattacac	20940
ctactctggc	tctataccct	acctagatgg	aaacttttat	ctcaaccaca	actttaagaa	21000
ggtgggcatt	acctttgact	ctctgttcag	ctggcctggc	aatgacggcc	cgcttaccce	21060
caacgagttt	gaaatataag	gctcagttga	cgggggaggtg	tacaacgttg	ccagctgtaa	21120
catgacaaa	gaactgtttc	tgtatacaat	gctagctaac	tacaacatg	gtctaccagg	21180
cttctatctc	ccagagagct	acaaggacgc	catgtactcc	ttctttagaa	acttccagcg	21240
catgagccgt	caggtgggtg	atgatactaa	atacaaggac	taccaacagg	tgggcatctc	21300
acaccaacac	acaaactctg	gattttgtgg	ctacottgcc	cccaccatgc	gcgaaggaca	21360
ggcctaccct	gctaaactcc	ccatctccgt	tataggcaag	accgcatgtg	acagcattac	21420
ccagaaaag	tttctttggc	atcgccacct	ttggcgcatc	ccatctccca	ctacttttat	21480
gtccatggcg	gcactcacag	acctgggcca	aaacctcttc	tacgcacact	ccgcccacgc	21540

gctagacatg	acttttgagg	tggatcccat	ggacgagccc	acccttcttt	atgttttgg	21600
tgaagtcttt	gacgtggtcc	gtgtgcacgc	gcgcacccgc	ggcgtcatcg	aaaccgtgta	21660
ctcgcgcacg	cccttctcgg	ccggcaacgc	cacaacataa	agaagcaacg	aacatacaaa	21720
acagctgcgc	cgatgggtcc	cagtgagcag	gaactgaaag	ccattgtcaa	agacttgggt	21780
tgtgggcgat	attttttggg	caactatgac	aagcgttttc	caggctttgt	ttctccacac	21840
aaagtcgcct	gcgcctatgt	caatacggcc	ggtcgcgaga	ctggggcggt	acaactggatg	21900
gcctttgctc	ggaaccgcga	ctcaaaaaca	tgtacctctt	ttgagccctt	tgctctttct	21960
gaccagcgac	tcaaggaggt	ttaccagttt	gagtacgaggt	caactcgcgc	cgttagcgcc	22020
attgctctct	cccccgagcg	ctgtataaac	ctggaaaagt	ccaccacaag	cgtacagggg	22080
cccaactcgg	cgcgctgtgg	actattctgc	tgcattgttc	tcacgccttt	tgccaactgg	22140
ccccaaaact	ccatggatca	caaccccacc	atgaacctta	ttaccggggt	accocactcc	22200
atgctcaaca	gtccccaggt	acagcccacc	ctcgctcgca	accaggaaca	gctctacacg	22260
ttcctggagc	gcccactgcgc	ctacttccgc	agccacagtg	cgcagattag	gagcgccact	22320
ttcttttggc	acttgtaaaa	catgtaaaaa	taactgtacta	gagacacttt	caataaagcg	22380
aaatgctttt	attttgtacac	tctcggttga	ttatttacc	ccacccttgc	cgtctcgccc	22440
gtttaaaaat	caaaagggtt	ctgcgcgcga	tgcgtatgcg	ccactggcag	ggagacggtt	22500
cgatactggt	gttttagtgt	ccacttaaac	tcaggacaaa	ccactcgggg	cagctcggtg	22560
aaagtttcac	ttccacaggt	gcgcacacat	accaacgcgt	ttagcaggtc	ggggccgcat	22620
atcttgaagt	cgcagttggg	gcctccgcgc	tgcgcgcgcg	agttgcgata	cagaggggtg	22680
cagcactgga	acactatcag	cgccgggtgg	tgcacgctgg	ccagcacgct	cttctcggaag	22740
atcagatccg	cgtccaggctc	ctccgcgttg	ctcagggcga	acggagtcaa	ctttggtagc	22800
tgctctccca	aaaaggggcg	gtgcccaggc	tttgagttgc	actcgcacgc	tagtggcatc	22860
aaaagggtgac	cgctgcgcgt	ctggggcgtta	ggatcacagc	ctctgctaaa	agacttgatc	22920
tgctttaaaag	ccactctgagc	ctttgcgcct	tcagagaaga	acatgcccgca	agacttgccg	22980
gaaaactgat	tgccgcgaca	ggccgcgtcg	tgcacgcagc	accttgcgtc	gggtgtggag	23040
atctgcacca	catctcgccc	ccaccgggtc	ttcacgatct	tggtcttgct	agactgctcc	23100
ttcagcgcgc	gctgcgcggt	ttcgctcgtc	acatccattt	caatcacgtg	ctccttattt	23160
atcataatgc	ttccgtgtag	acacttaagc	tcgccttcga	tctcagcgca	gggtgtgacg	23220
cacaacgcgc	agcccggtgg	ctcgtgatgc	ttgtagggtca	ccctctgaaa	cgacttcagg	23280
tcagcctcga	ggaatcgccc	catcatcgtc	acaaaggtct	tgttgtcgtg	gaaggtcagc	23340
tgcaaccgcg	ggtgctcttc	gttcagccag	gtcttgcata	cgcgcgccag	agcttccact	23400
tggtcagcga	gtagtttgaa	gttcgccttt	agatcgttat	ccacgtggta	cttgtccact	23460
agcgcgcgcg	cagctcccat	gcctttcttc	cacgcagaca	cgatcggcac	acctcagcgg	23520
ttcatcaccg	taatttcaat	ttccgcttcg	ctgggctctt	ctcttctctc	ttcgctccgc	23580
atcaacgcgc	ccactgggtc	gtcttcaatc	agccgcgcga	ctgtgcgctt	acctcctttg	23640
ccatgcttga	ttagcacccg	tggtgtgtgt	aaaccaccca	tttgtagcgc	caatcttctt	23700
ctttcttctc	cgtctccac	gattacctct	gggtgatggc	ggcgtcggg	ctttgggagaa	23760
gggcgctctc	ttttttcttt	ggcgcgaatg	gccaaatccg	ccgcgcaggt	cgatggccgc	23820
gggctgggtg	tgccgcgcac	cagcgcgtct	tgtgatgagt	cttccctgtc	ctcggaactg	23880
atagccgcgc	tcacccgctt	ttttgggggc	gcccgccgag	cgcggccgga	cggggacggg	23940
gacgaacagt	ctctccatgt	tgggggagct	cgcgcgcac	cgcgtccgcg	ctcgggggtg	24000
gtttgcgctc	gtctctcttc	ccgactggcc	atttcttctc	ccctataggca	gaaaagatc	24060
gaggtgtcag	tcgagaagaa	ggacagccca	accgcctctc	ctgagttcgc	caccacgcgc	24120
tcaccagatg	ccgccaacgc	gcctaccacc	ttcccgctgc	aggcaccccc	gcttgagagc	24180
gaggaagtga	ttatcgcgca	ggaccacaggt	tttgtaaagc	aagacgacga	ggaccgctca	24240
gtaccaacag	aggataaaaa	gcaagaccag	gacaacgcag	aggcaaacga	ggaacagctc	24300
ggggcggggg	agcaaaaggca	tgccgacatc	ctagatgtgg	ggagcagcgt	gctgttgaa	24360
catctgcagc	gccagtgccg	caattatctg	gacgcgttgc	aagagcgcag	cgttgtgccc	24420
ctcgcgatag	cggagtctag	ctttgctcac	gaaccgcgcg	tattctcacc	gcgcgtaccc	24480
ccccaaacgc	agcaaaacgc	acatacgag	ccccaccgcg	gcctcaactt	ctaccccgtg	24540
ttbgcgtgtc	cagaggtgtct	tgccacctat	cacattcttt	tccaaaactg	caagatcccc	24600
ctatctgcgc	gtgcccaacg	cagccgagcg	gacagacgag	tggtcttgcc	gcagggcgct	24660
gtcatacctg	atatcgcttc	gctcaacgaa	gtgccaaaaa	tctttgaggg	tcttggaacg	24720
gacgagaagc	gcgcgcgcaa	cgtcttcgaa	caggaaaaac	gcgaaaaatg	aagtcactct	24780
ggagtgtttg	tggaaactga	gggtgacaac	gcgcgcctag	cgtactaaaa	acgcagcatc	24840

gaggtcacc	acttttgcta	cccgccactt	aacctacc	ccaaggtcat	gagcacagtc	24900
atgagtgagc	tgatcgtg	ccgtgcgcag	ccctcggaga	gggaagcaaa	ttgcgaagaa	24960
caaacagagc	agggccctacc	cgcagttggc	gaacgagcgc	tgcgcgctgc	gcttcaaacg	25020
cgcgagccgt	cgacatctga	ggagcgacgc	aaactaatga	tggccgcagc	gctcgttacc	25080
gtggagcttg	agtcatctga	cggtttcttt	gctgacccgc	agatcgagcg	caagctagag	25140
gaacatctgc	gcacaacctt	tcgacagggc	tacgtaccgc	agggcctgcaa	gatctccaa	25200
gtggagcttc	gcacaacctt	ctcctacctt	ggaattttgc	acgaaaaacc	ccttggggcaa	25260
aaagtgtctc	attccacgct	caagggcgag	gcgcgcgcgc	actacgtccg	cgactcggtt	25320
tacttatttc	tatgcctacac	ctggcagacg	gccattggcg	tttggcagca	gagcttgag	25380
gagtgcaacc	tcaaggagct	gcagaaactc	ctaaagcaaa	acttgaagga	cttatggagc	25440
gccttcaacg	agcgtccgt	ggccgcgcac	ctggcgggca	tcattttccc	cgaaacgctg	25500
ctttaaaccg	tcctacaggg	tctgcacagc	ttcaccagtc	aaagcatgtt	gcagaacttt	25560
aggaacttta	tccatagagc	ctcaggaatt	ttgcgcgcgc	cctgctgtgc	acttccagc	25620
gactttgtgc	ccattaaagta	ccgcgaatgc	cttcgcgcgc	tttggggcca	ctgcctacctt	25680
ctgcagctag	ccaactacct	tgccctaccac	cttgacataa	tggaaagacgt	gagcggtgac	25740
ggctctactg	agtgctactg	tcgctgcacac	ctatgcaccc	cgccaccgctc	cctgggtttgc	25800
aatctgcagc	tgcttaacga	aagtcaaat	atcggtacct	ttgagctgca	gagtcctctg	25860
ctctgacaaa	agtcgcgcgc	tcgggggttg	aaactcactc	cggggctgtg	gagtcctctg	25920
tactctcgca	aatttctacc	tgaggactac	cacgcccacg	agattaggtt	ctacgaagac	25980
caatcccgcc	cgccaaatgc	ggagcttacc	gcctgcgtca	ttaccacagg	ccactctctt	26040
ggccactcgc	agccatcaaa	caagagccgc	caagagtttc	tgctacgaaa	ggggccgggg	26100
gtttacttgg	accccagctc	cgccgagggc	ctcaacccaa	ttcccccgcc	cgccgagccg	26160
tatcagcagc	agcccggggc	ccttgcttcc	caggatggca	ccccaaaaga	cttgctgagc	26220
gcgcgcgcga	cccagcgagc	aggaggaata	ctgggacagt	caggcagagc	aggttttgga	26280
cgaggagggag	gagcacatga	tggaaagact	ggagagccta	cttcgagga	cttcgaggtt	26340
cgaaaggggt	tcagacgaaa	caccgtcac	ctcggctgca	ttccctcgc	cgccgcccga	26400
gaatcgcgca	accggttcca	gcattggctac	aacctccgct	ctcagggcgc	cgccggcact	26460
gccgcttcgc	cgaccacaac	gtatagtgga	caccactgga	accagggcgc	gtaagtccaa	26520
gcagccgcgc	ccgttagccgc	aagagcaaca	acagcgccaa	ggctaccgct	catggccgcg	26580
gcacagaagc	gcattagtgt	cttgcttgca	agactgtggg	ggcaacatct	ccttcgcccgc	26640
ccgctttctt	ctctaccatc	acggcgctgg	cttccccgc	aacatcctgc	attactacg	26700
tcctctctac	agccctactc	gcaccggcgc	cagcggcgag	ggcagcaaca	gcagcggcca	26760
ccagagaagc	aaggcgaccg	gatagcaaga	ctctgacaaa	gcccagaaga	tcacagcg	26820
cgccagcagc	aggagggagg	gcgctgcgtc	tggcgcccaa	cgaacccgta	tcgaccgcgc	26880
agcttagaaa	caggattttt	ccactctgt	atgtatatt	tcacagagc	agggggcaag	26940
aaacagagct	gaaaaataaa	aaagggtctc	tgcgatccct	caccgcgagc	tgctgtatc	27000
acaaaacgca	agatcagctt	cgccgcacgc	tggaaagcgc	ggaggctctc	ttcagtaaat	27060
actcgcgcct	gactcttaag	gactagtctc	gcgcctttc	tcaaatttaa	gcgcgaaaa	27120
tacgtcatct	ccagcgccca	caccgcgcgc	cagcactgt	gctcagcgc	attatgagca	27180
aggaatctcc	cacgcccact	atgtggagtt	accagccaca	aatgggactt	gcgcgtggag	27240
gttcccaagc	ctactcaacc	caataaaact	acatgagcc	gggaccctac	actgatctcc	27300
ccacactcgc	aatccgcgc	gcagaaacc	gaattctctt	gggaacagcg	gctattacca	27360
gtcccctcc	caccactgtg	gtacttccca	gagacgcgca	ggccgaagt	taccaggaaa	27420
actcaggggc	gcagcttgcc	ggcggttttc	gtcacagggt	cgcggtgcgc	ggcgaggata	27480
taactcaact	gcacatcaga	ggcgaggtta	gtcagctcaa	gcacagatcg	gctagctctc	27540
cgcttggtct	ccgtcccgag	gggacatttc	agatcgcggc	cgccggcgct	ccttcattca	27600
cgctcgtcta	accttgacga	actctgcaga	cctcgtcttc	tgagcgcgcg	cttgaggagca	27660
ttggaaactc	gcaatttatt	gaggagtttg	tgccatcggt	ctactttaac	ccctctctgc	27720
gaactccgcg	ccactatccg	gatcaattta	ttcctaaact	tga-cgcggtta	aaggatcctg	27780
cggaagcgcta	cgactgaagt	tttaagtggag	aggcagagca	actgcgcgct	aaacacactg	27840
tcacactgtc	cgcccaacaag	tgctttggcc	gcgactccgc	tgagttttgc	tactttgaat	27900
tgcccgagga	tcattctcag	ggcccgcgcc	acggcgctgc	gcttaccgcc	cagggaagagc	27960
ttgcgcgtag	cctgattcgg	gagtttacc	agcgccccc	ctagtttgag	cgggagagagc	28020
gacctctgtg	tctcactgtg	attttgcaact	gtcctaact	tggattacat	caagatcttt	28080

gttgccatct	ctgtgctgag	tataataaat	acagaaatta	aaatatactg	gggctectat	28200
cgccatctcg	taaacgccac	cgtcttcacc	cgcccaagca	aacccaagcg	aacctactct	28205
ggatcatttta	acatctctcc	ctctgtgatt	tacacacagt	tcaaccagga	cgaggtaggt	28320
ctacagagaa	acctctccga	gctcagctac	tcacatcgaa	aaaaccacc	ctctcttacc	28380
gtgcgggaac	ctgacagtcg	gtcaccggcc	gctgcaccac	acctcagccc	tgacggtaaa	28440
ccagactttt	tcgggcagca	ctcacaatac	ctctgtttacc	agaacaggag	gtggacttag	28500
aaaaaccotta	gggtatttag	ccaaaggcgc	agctactgtg	gggtttatga	acaattcaag	28560
caactctcag	ggctatttcta	attcagggtt	ctctagaatc	gggtgtgggg	ttattctctg	28620
ctctgtgatt	ctctcttattc	ttatactaac	gctctctctg	ctaaggctcg	ccgctctctg	28680
gtgcacacatt	tgcatattatt	gtcagctttt	taaacctggy	ggctgcacc	caagatgatt	28740
aggtacataaa	tcttaggttt	actcaccctt	gcytcagccc	acggtaccac	ccaaaagggt	28800
gatttttaagg	agccagcctg	taattgtaca	ttcgacagct	aagctaata	gtgcaccact	28860
cttataaaat	gcaccacaga	acatgaaaag	ctgcttattc	gccacaaaa	caaaattgpc	28920
aagatgatcgt	tttatgctat	ttggcgacca	ggtagacacta	cagagataaa	tgtagcagtt	28980
ttccagggtta	aaagtcatata	aacttttatg	tatacttttc	cattttatga	aatgtgccc	29040
attaccatgt	acatgagcaa	acagataaag	ttgtggcccc	cacaaaaatt	tgtaggaata	29100
actggtcactt	tctgtgcac	tgctatgcta	attacagtgc	tcgctttggt	ctgtacccta	29160
ctctataatta	aatacaaaag	cagacgcagc	tttattgagg	aaagaaaaat	gccttaattt	29220
actaagttac	aaagctaatg	tcaccactaa	ctgcttttgc	cgctgtctgc	aaaaacaaat	29280
caaaaaggtta	gcattataat	tagaataagg	tttaaacccc	ccggtcattt	ctctgtcaat	29340
accattcccc	tgaaacaattg	actctatgtg	ggatgatgctc	cagcgctaca	caattgtgag	29400
caggctctctc	ggatgtcagc	atctgacttt	ggccagcacc	tgctccgcgg	attgttccca	29460
gtccacacac	agcgaccacc	ctcaacagag	atgaccaaca	caaccaacgc	ggccgcgcgt	29520
accggactta	catctaccac	aaatacacc	caagtttctg	ctttgtgcaa	taactgggat	29580
aacttgggca	ttgtgtggtt	ctccatagcg	cttatgtttg	tatgccttat	tattatgtgg	29640
ctcatctgct	gcctaaagcg	caaacgcgcc	cgaccaccca	tcctatagtc	catcatgtgt	29700
ctacacccaa	acaatgatgg	aatccataga	ttggacggag	tgaaacacat	gtctttttct	29760
cttacagtat	gattaaatga	gacatgattc	ctcgagtttt	tataattactg	acctttgtgt	29820
cgctttttttg	tgctgtctcc	acattggctg	cggtttctca	catcgaagta	gactgcattc	29880
cagcctctcac	agtcctattg	ctttacggat	ttgtccacct	cagctctcat	tgacgcctca	29940
tcactgtgggt	catcgccctt	atccagtgc	ttgactgggt	ctgtgtgccc	tttgcatatc	30000
tcagacacca	tcctcagctac	aggagacagg	ctatagctga	gctttctaga	attctttaat	30060
tatgaatttt	actgtgactt	ttctgctgat	tatttgaccc	ctatctgcgt	tttgttcccc	30120
gacctccaaag	ctcacaagac	atatatcatg	cagattcact	cgatatagga	atatcccaag	30180
ttgctacaat	gaaaaaagcg	atctttccga	agcctgggta	tatgcaatca	ttctctgtat	30240
gggtgtctctg	agtaccatct	tagccctagc	tatatatccc	taccttgaca	ttgctgggaa	30300
acgaatagat	gccatgaacc	accacaactt	ccccgcgcgc	gctatgcttc	actgcgaaca	30360
agttgtgtgc	ggcggtcttg	tcaccgccaa	tcagcctcgc	ccacttctc	ccacccccac	30420
tgaaatcagc	taactttaatc	taacaggagg	agatgactga	acactagat	ctagaaaagg	30480
acggaaatcat	tcacagagcag	cgctgctgag	aaagacgcag	ggcgagcgcc	gagcaacagc	30540
gcataatata	agagctccaa	gacatgggta	acttgaccca	gtgcaaaagg	ggatcttttt	30600
gtctgtgtaa	cgaggccaaa	gtcaactacg	acagtataac	accggacac	cgctcttagct	30660
acaagttgcc	aaccagcggt	cagaatttgg	tggtctatgt	gggagaaaag	cccatatcca	30720
taactcagca	ctcggtagaa	accgaaggct	gcattcaact	accttgtoaa	ggactctgag	30780
attctctgac	ctctattaag	acctgtgtgc	gtctcaaga	tccttatccc	tttaactaat	30840
aaaaaaaatat	ataaaagcat	caacttacta	aaatcagtta	gcaattttct	gtccagttta	30900
ttccagagca	ctctcttgcc	ctctctccag	ctctgttgat	gcagcttctc	ctcgctgtca	30960
aactttctcc	acaaactctaa	tggaatgca	gttctctctc	gttctgtccc	atccgcaccc	31020
actatcttca	ttgtgttgca	gatgaagcgc	gcaagacggt	ctgaagatcc	cttcaacccc	31080
gtgtatccat	atgacacgga	aacggttctc	ccaactgtgc	ttctttctac	tcctcccttt	31140
gtatccccca	atgggttttca	agagagtccc	ccctgggtac	tccttttgcc	cttatccgaa	31200
ctctagtta	ctcccaatgg	catgcttgcc	ctcaaaatgg	gcacagggct	ctctctggag	31260
ggggccggca	accttaccctc	ccaaaaatga	accaactctg	gcccaacctc	caaaaaaac	31320
aggtcaaaac	taaacctcga	aatatctgca	ccctcagata	ttacctcaag	agcccttaact	31380
gtggctgcgc	ccgcacactct	aatggtgcgc	ggcaacacac	tcacctagca	atcacaggcc	31440

ccgcataaccg	tgacagcactc	caaaccttagc	attgccacccc	aaggacccccc	cacagtgtca	31500
gaaggaaagcg	tagccctgcga	aacatcaggcg	cccctcaccca	ccaccgatagc	cagtagaccctt	31560
actatcacctg	ctcacacccc	tctaactact	gccacttgta	gcttggggcat	tgacttggaaa	31620
gagcccatctt	atacacaaaa	tgaaaanaacta	ggactaaagt	acgggggctcc	tttgcattgta	31680
acagacgacc	taaacactttt	gacgtagca	actggtccag	gtgtgactat	taataataact	31740
tccttgcmaa	ctaaagttac	tgagagcctg	ggtttttgatt	cacaaggcaa	taagcaactt	31800
aatgtgagcag	gaggactaag	gattgattct	caaaacagac	gccttatact	tgatgttagt	31860
tatccgtttg	atgtctaaaa	ccaactaaat	ctaagacag	gacaggggccc	tccttttata	31920
aactcagccc	acaacttggg	tattaactac	aacaaaagccc	tttacttgtt	tacagcttca	31980
aaacaattcca	aaaagcttga	ggttaaacta	agcactgcc	aggggttgat	gtttgacgct	32040
acagccatag	ccattaatgc	aggagatggg	cttgaatttg	gttcacctaa	tgccacaaac	32100
acaaactccc	ctaaaaacaa	aattggccat	ggcctagaat	ttgattcaaa	caaggctatg	32160
gtctctaacc	tgaggactgg	ccttagtttt	gacagcacag	gtgccat tac	agtgagaaac	32220
aaaaataatg	ataagctaac	tttgggacc	acaccagctc	catctcctaa	ctgtagacta	32280
aatgcagaga	aagatgctaa	actcactttg	gtcttaacaa	aatgtggcag	tcaaatactt	32340
gctacagttt	cagttttggc	tgtaaaggcc	agtttggctc	caatatctgg	aacagttcaa	32400
agtgctcctc	ttattataag	atttgacgaa	aatggagctc	tactaaacaa	ttcccttcctg	32460
gacccagaat	atttgaactt	tagaaaatgga	gatcttactg	aaggcacagc	ctatcaaaa	32520
gctgtgggat	ttatgcctaa	cctatcagct	tatccaaaa	ctcaccgtta	aaactgcmaa	32580
agtaacattg	tcagtaaatg	ttacttaaac	ggagacaaaa	ctaaacctgt	aaactcaaac	32640
attacactaa	acggtcacaca	ggaacacagg	gacacaaact	caagtgcata	ctctatgtca	32700
ttttcatggg	actgtgtctg	ccacaactac	attaatgaaa	tatttggcct	atccctttac	32760
actttttcat	acttggccca	agaataaaga	atcgttttgg	ttatgtttca	acgtgtttat	32820
ttttcaattg	cagaaaattt	caagtcaatt	ttcattcagt	agtatagccc	caccaccaca	32880
tagctctatac	agataccagt	accttaatac	aactcaacaa	accctagtat	tcaacctgcc	32940
acctccctcc	caacacacag	agtaacacag	cccttctccc	cggtgtgccc	taaaaagcat	33000
catatcatgg	gtaacagaca	tattcttagg	tgttatattc	cacacgggtt	ccgtgtcgag	33060
caaacgcctca	tcagtgtatg	taataaactc	cccgggcagc	tcacttaagt	tcattgtcgt	33120
gtccagctgc	tgagccacag	gctgctgtcc	aacttgcggt	tgcttaacgg	gcgggtgaagg	33180
agaagtccac	gctacatagg	gggtagagct	ataatcgtgc	atcaggatag	ggcgggtggg	33240
ctgcagcagc	gcgcgaataa	actgctgcgc	cgccgcctcc	gtcctgcagc	aatacaaat	33300
ggcagtggtc	tctcagcgca	tgattgcgac	cgccgcgcag	ataaggcgcc	ttgtctctcc	33360
ggccagcgag	cgacccctga	tctcacttaa	atcagcacag	taactgcagc	acagcacacc	33420
aaatattgtt	aaaatccccc	agtgcaaggc	gctgtatcca	aagctcatgg	cggggaccac	33480
agaaacccagc	tgccctcatc	accacaagcg	caggtagatt	aagtggcgac	ccctcataaa	33540
caagctggac	ataaacatta	ctctttttgg	catgtttgaa	ttcaccacct	cccggtacca	33600
tataaacctc	tgattaaaca	tgccgcctac	caccacatc	ctaaaccagc	tgcccaaaa	33660
ctgcccgcgc	gctatcatct	gcagggaacc	gggaactgaa	caatgacagt	ggagagccca	33720
ggactgcata	ccatggatca	tcattgctgt	catgatata	atgttggcac	aacacaggca	33780
cacgtgtgca	cacttctctca	ggattacaag	ctctctccgc	gttagaacca	tatccacagg	33840
aaacaacctat	tcttgataca	ggtaaatcc	caactgcag	ggaagaccc	gcagtaact	33900
ccggtgtgct	attgtcaaa	tgttacattc	gggcagcagc	ggatgatctc	ccagttaggt	33960
agccggggct	tctgttccaa	aaggaggtag	acgatcccta	ctgtacggag	tgccgcgaga	34020
caaccagatg	ctgtgtggctc	gtatgtctat	gccaaaatgga	acgcccggagc	tagtctattt	34080
tctctgaagca	aaaccagggtg	ggggcgtgac	aaacagatct	gcgtctccgg	ttctgcgctt	34140
tagatgcctc	ttgtgtagtag	ttgtagtata	tcacctctct	ccatgataca	ccaggtgcctc	34200
tggtctcggg	ttctatgtaa	actcctctat	ggccgctgct	cctgataaca	tcacacagcc	34260
cagaataagc	cacacccagc	caacctacac	attctgtctg	cgagtcacac	acgggagagag	34320
cggaagagcg	tggaagaacc	atgttttttt	tttttatcca	aaagattatc	caaaacctca	34380
aaatggagat	ctatttaagt	aacgcctctc	cctccggtgg	cggtgtcaaa	ctgtcacagcc	34440
aaagacacga	taattgcatct	tgtaaagtgt	tgacacatgg	cttccaaaaa	gcaaacggcc	34500
ctcagctcca	agtgagccta	aaggctaaac	ccttcagggt	ccctctctct	tataaacatt	34560
ccagccacctt	caacctatgcc	caataattct	tcatctcgc	accttctcaa	tatatctcta	34620
agcaaatccc	gaattattaag	tcggcgcat	gtcaaaaatc	gtccacagag	gcctctcacc	34680
tcagcctcca	agcagcgat	catgatgtca	aaaattcagg	ttctctcacag	acctgtataa	34740

gattcaaaaag	cgggaacatta	acaaaaatac	cgcgaccccg	taggtccctt	cgcaggggcca	34800
gctgaacata	actcgtgcagg	tctgcacgga	ccagcgcggc	cacttccccg	ccaggaaacct	34860
tgacaaaaga	accocacactg	attatgacac	gcatactcgg	agctatgcta	accagcgtg	34920
ccccgatgta	agctttgttg	catggggcgg	gatataaat	gcaagggtgc	gctcaaaaaa	34980
tcaggcaaaag	cctcggcgaa	aaaagaaagc	acatcgtagt	catgctcatg	cagataaagg	35040
caggtaagct	cgggaaccac	cacagaaaaa	gacaccattt	ttctctaaa	catgtctcgg	35100
ggtttctgca	taaacacaaa	ataaaataac	aaaaaacat	ttaaacatta	gaagcctgtc	35160
ttacaacagg	aaaaacaacc	cttataagca	taagacggag	tacggccatg	cggcggtgac	35220
cttaaaaaaa	ctgggtcacg	tgattaaaaa	gcacccacga	cagctcctcg	gtcatgtccg	35280
gagtcataat	gtaagactcg	gtaaacacat	caggttgatt	catcggtcag	tgctaaaaag	35340
cgaccgaaat	agcccggggg	aatacatacc	cgcaggcgta	gagacaacat	tacagcccc	35400
ataggaggta	taacaaaatt	aataggagag	aaaaacacat	aaacacctga	aaaacccctc	35460
tgcttaggca	aaatagcacc	ctcccgtccc	agaacaacat	acagcgcttc	acagcgggag	35520
cctaaccagtc	agccttacc	gtaaaaaaga	aaacctatta	aaaaaacacc	actcgacagc	35580
gcaccagctc	aatcagtcac	agtgtaaaaa	agggccaagt	ccagagcgag	tatatatgag	35640
actaaaaaat	gacgtaacgg	ttaaagtcca	caaaaaacac	ccagaaaaac	gcaacgggag	35700
ctacgcccag	aaacgaaagc	caaaaaaccc	acaacttccc	caaatcgtca	cttcctgttt	35760
cccacgttac	gtactctccc	atttttaagaa	aactacaatt	cccaacacat	acaaagtact	35820
cgcgccataa	actcactgta	cgcgcctcgt	tcccacgccc	cgcgcacagt	caaaaactcc	35880
acccccatc	tatcatattg	gcttcaatcc	aaaataaggt	atatatttga	tgatg	35935

<210> 10

<211> 5965

<212> DNA

<213> Artificial Sequence

<220>

<223> NSsuboptmut

<400> 10

gtccacatgg	cccccatcac	cgcctacagc	cagcagacca	ggggcctgct	gggtgcatac	60
ataccacagcc	tgacggcgagc	cgacaagaac	caggttgagg	gagaggtgca	ggtggtgagc	120
accgctaccc	agagcttccc	ggccacctgc	gtgaacggcg	tggtctggag	cgtgtaccac	180
ggagccggaa	cgaaagacct	ggccggagcc	aaggggcccta	tcaccagatg	gtacacccat	240
gtggatacgg	attctgggtggg	ctggcaggcc	cctccgggag	ccaggagcct	gacacctgtg	300
acctgtggaa	cagcgcacct	gtacctgggt	acacgcacag	ccgatgtgat	cccgtgagg	360
cgcaggggcg	attctcggcg	aagcctgctg	agccctaggc	cctgtgagcta	cctgaagggc	420
agcagcggag	gaccctcgtc	gtgtcctctt	ggccatgcgc	tgggcatatt	tgcgctgccc	480
gtgtgtacca	ggggcgtggc	caaaagccgtg	gattttgtgc	cctgtgaaag	catgagagac	540
atccatcgca	gcctctgtgt	cacgcgaac	agctctcccc	ctgcgctgccc	ccaatcattc	600
caggttgctc	acctgcagcg	ccctacccgga	tctggcaaga	gcaccaaggt	gcocgtgccc	660
tacgcgcgtc	agggctacaaa	ggctgctggg	ctgaacccca	gcgtggccgc	taccctgggg	720
ttcggcgctt	acatgagacaa	ggcccatggc	atcgacccca	acatccgcac	agggcgtggc	780
accaacacca	cgggagctcc	cgtgacctac	agcacttacg	ccaagttccc	ggcggatgga	840
ggctgcagcg	gaggagacct	cgacatcacc	atctgcgagc	agtgccacag	caccgacagc	900
accacacatc	tgggcattgg	caccgtgtgt	gatcaggccg	aaacagctgg	agccagctgt	960
gtgggtgctg	ccacagctac	cctccttggt	agcgtgaccg	tgccccatcc	caatctcgag	1020
gaggtggccc	tgagcaaacac	agggcgagtc	cctctctacg	gcaaggccat	ccccatcgag	1080
ggcatccggc	gaggcaggca	cctgatcttc	tgccacagca	agaagaagtg	cgacgagctg	1140
gctgcacagg	tgagcgactc	gggcataaac	gcctgtggct	actacagggg	cctggagctg	1200
tcagtgatcc	ccaccatcgg	cgatgtgggt	gtggtggcca	ccgacgcccc	gatgacaagg	1260
tacacggggg	actctgacag	cgtgatcgac	tgcaaacacct	gcgtgaccca	gaccgtggag	1320
ttcagcctgg	accccacctt	accactcgaa	accacacacc	tgccctcagg	tgctgtggag	1380
agggaccaga	ggcgcggcag	caccggaaag	ggcaggcgcg	gaatttatcg	cttctgtgac	1440
cctggcgaaa	ggccctctgg	catgtctgac	agcagcgtgc	tgctcgagtg	ctacgacgct	1500

ggctgcgctt	ggtagcagct	gacacccgct	gaaccacgag	tgcgcctgcg	cgttatctctg	1560
aatcccccctg	gccgcgccgt	gtgtcaggac	cacctggagt	tctgggagag	cgtgttcaca	1620
ggactgaccc	acatcgacgc	ccatttcctg	agccagacca	agcaggctgg	cgcacacttc	1680
ccctatctg	tgccgtatca	gyccaccgtg	tgtgtatcgg	cccaagctcc	acctctctca	1740
tgggaccaga	tgtggaagtg	cctgatccgc	ctgaagccga	ccctgcacgg	ccctaccctc	1800
ctgctgtacc	gccctgggagc	cgtgcagaa	gaggtgaacc	tgaccacacc	catcaccaga	1860
tacatcatgg	cctgcctagc	cgtgatctgc	gaagtgtgtga	ccagcacctg	gggtcgtgtg	1920
ggaggcgtgc	tggccgctct	ggctgcctac	tgcctgacca	ccggaaagct	gggtgatcgtg	1980
ggacgcgtca	tccctgagcgg	aaggcccgct	atcgtgcctg	atcgcgagtt	ccgtgtaccag	2040
gagttcgagc	agatgggagca	gtgtgccacc	cacctgcctc	acatcgagca	ggggatgcag	2100
ctggccgaac	agttcaagca	gaagccctcg	ggcctgtctg	agacagccac	caaacaggcc	2160
gaagctgccg	ctcccgtgtg	ggaaagcaag	tggagggcc	tggagacctt	ctgggctaa	2220
cacatgttga	acttcattctc	tggcatccag	tacctggccg	gactgagcac	cgtcctgggc	2280
aaccccgccta	tcgccagcct	gatggccttc	accgttagca	tcacctctcc	ccgtaccacc	2340
cagagcaccc	tgcctttcaa	cattctgggc	ggatgggtgg	ccgctcagct	ggcccctctc	2400
tcagctgctt	ctgcctttgt	ggcgctggc	attgccggag	ccgctgtggg	cagcatgtgc	2460
ctgggcaaa	tgcctgttga	tattctggct	ggctatggcg	ctggcgtggc	cggaggccctg	2520
cgtgccttca	aggtgatgag	cggagagatg	cccagcaccg	aggacctggt	gaacctgtctg	2580
cttgccattc	tgaagccctg	agccctgggtg	gtggcgctgtg	tgtgtgcttc	catcttgagc	2640
cgccattgtg	gaccocggaga	ggcgctgtgtg	cagtggaatga	accgcctgat	cgcctctgcc	2700
ttctcgggaa	accacgctgag	ccctaccacc	tacgtgctctg	agagcgacgc	cgtctggcag	2760
gtgaccaca	tccctgagcag	cctgaccattc	accagctgc	tgaagcgctc	gcaccagtgtg	2820
atacaagag	actgcagcac	accctgcagc	ggaaagctgc	tgagggaactg	gtgggactgtg	2880
atctgcaccg	tgcctgaccga	cttcaagacc	tggctgcaga	gcaagctgct	gcccacaactg	2940
ctggcgctgc	ccctctcttc	atgccagcgc	ggatacaagg	cgctgtggag	ggcgatgtgc	3000
atcatcgaga	ccacctgtcc	ctggcgagcc	cagatcacag	gccacgtgaa	gaaccgcagc	3060
atgcgcatac	tgggcctcaa	gaacctgcagc	aacacctggc	acggcacctt	ccctctaac	3120
gccataacca	ccggaccctg	cacaccacgc	cctgtctcca	actacagcag	ggccctgtgtg	3180
agggcgctgc	ccgaggagta	cgtggagggtg	accagggttg	gagacttcca	ctacgtgacc	3240
ggatgacca	ccgaccaactg	gaagtgtccc	tgtcaggtgc	ccgctcccca	attttttacc	3300
gaagtgcgatg	gcgtgcgcct	gcactgcctat	gccctgcct	tgaggccctc	gctgcgcgaa	3360
gaagtgcact	tccaggtggg	cctgaaccag	tacctggtgc	gcagcagctg	gccctgcagc	3420
ccctgagccc	atgtggccgt	gctgaccagc	tcctctccaa	accaccagca	catcacagcc	3480
gaaccgccta	aaaggcgctc	ggccaggggc	tcctctccaa	gcccgtccctc	aagcagcctc	3540
agccacctgt	ctgctccacc	cctgaaggcc	acctgcaaca	cccaccagct	gagcccagac	3600
gccgacctga	tgaaggccaa	cctgctgtgg	cgccaggaga	tgggcggcaa	catcacccgc	3660
gtggaggagc	agaacaaggt	ggtgtgtgctg	gacagctctg	accccctgcg	cgccgagggag	3720
gacgagcgcg	aggtgagcgt	gcccgccag	atctctgcga	agagcaagaa	gttcccctgc	3780
gccatgccca	tcctgggtcag	acctgattac	aacctctccc	tgtctgagag	ctgggaagac	3840
ctgatctacg	tgcctccagt	ggctcatggc	tgtctctctg	ctcccattaa	agccctctcc	3900
atctccacctc	ctaggcgcaa	aaggacccgtg	gtgctgacag	aaagcagcgt	gagctctgct	3960
ctggcggaac	tggccaccaa	gacctttggc	agcagcgaga	gctctgcgtg	ggacagcgga	4020
acagccacgc	ctctgctgtg	ccaggccagc	gacgacggcg	ataaggcgag	cgatgtgag	4080
agctatagca	gcattgcctcc	cctggaaggc	gaacctggcg	atcccgatct	gagcgtatgg	4140
agctggagca	ccgtgagcga	agaggccagc	gaggacgtgg	tgtgttgacg	catgagctac	4200
acctggacag	gcgctctgat	cacacctgcg	gctgcccagag	agagcaagct	gcccactaac	4260
gcccttgagca	acagcctgct	gaggcaccac	aacatggtgt	acggccaccac	cagcaggtct	4320
gccggagtag	ggcagaagaa	ggtagcttc	gaccgctctg	aggtgtgtga	gcagcaatc	4380
cgctgtgtgc	tgaaggaagt	gaaggccaa	gccagcaccg	tgaaggccaa	gctgtgagc	4440
gtggaggagg	ccctgcaagct	gaccccccc	cacagcgcca	agagcaagt	cgctacagcg	4500
gccaaaggagc	tgcgcaacct	gagcagcaag	gccgtgaacc	acatccacag	cgtgtggaag	4560
gacctgtctgg	aggacacgtg	gacccccatc	gacacaccca	tcatggccaa	gaacgaggtg	4620
ttctgcgtgc	agccccagaa	ggggcgccgc	aagcccgtct	gcctgactgt	gttccccgat	4680
ctggcgctgc	cgctgtgcga	gaagatggcc	ctgtacgacg	tgtgagcac	cctgcctcag	4740
gtggtgatgg	gctcaagcta	cggctccag	tacagccctg	gcagcgcgt	ggagttctctg	4800

gtgaacaccc	ggaagagcaa	gaagaacccc	atgggcttca	gtacgacac	acgctgcttc	4860
gacagcagcc	tgaccgagaa	cgacatccgc	gtggaggaga	gcattacaca	gtgctcgagc	4920
ctggccccct	agggcaggga	ggccatcaag	agcctgaccg	agcgcttgta	catcgagggc	4980
ccctctacca	acagcaaggg	acagaaactgc	ggatacaggc	gctgtagggc	ctctggcgctg	5040
ctgacaacca	gctgtggcaa	cacctgacc	tgctacctga	agggccagcg	tgccgtgtgc	5100
gtgcccgaag	tgacgactg	caccatgctg	gtgaacgcgc	ctggcctggg	ggatgtttgt	5160
gaaagcgctg	gcaccaggga	agatgtgccc	agcctgcgcg	tgctccacga	ggccatgacc	5220
aggtactctg	ccccctccgg	agacccccct	cagcccgaa	acgacctgga	gctgatcacc	5280
agctgtctaa	gcaacgtgag	cgctggctcac	gacgccagcg	gaaagcgctg	gtactacctg	5340
acacggatgc	ccaccacccc	tctggctcgc	gctgcttggg	aaaccgctcg	ccatacacc	5400
gtgaacagct	ggctggggcaa	catcatcatg	tacggcccta	ccctgtgggc	tcgatgata	5460
ctgatgacct	acttcttccg	catctgtgct	gctcaggagc	agctggagaa	ggccctggag	5520
tgccagattt	acggcgctct	ctacagcctc	gagccctctg	acctgcccc	aatacatcga	5580
cgccctcacg	gctctgtctg	cttcagcctg	cacagctaca	gcccctggca	aatacatcgc	5640
gtggccagct	gtctgcgcaa	actggcgctg	ccctctctgc	gcgtgtggag	gcatagggct	5700
aggagcgctg	gggttaggct	gctgagccag	ggaggcaggg	ccgtacctct	tggaaagtac	5760
ctgttcaact	gggcgctgaa	gaccaagctg	aagctgaccc	ctatccctgc	cgctacccgc	5820
ctggagctga	gggatgtggt	cgctgtctgc	tacagcgag	gcgacatcta	ccacagcctg	5880
tctcgcgtc	gcctctcgctg	gttcatgctg	tgctgtctgc	tgctgagcgt	gggcgtgggc	5940
atctacctgc	tgcccaaccg	ctaaa				5965

<210> 11

<211> 5965

<212> DNA

<213> Artificial Sequence

<220>

<223> Chimeric NSsuboptmut

<400> 11

gccaccatgg	cccccatcac	cgctacagc	cagcagacc	gcggcctgct	gggctgcata	60
atcaccagcc	tgaccggcgc	cgacaagaac	caggtggagg	gcgaggtgca	gggtgtgagc	120
accgccaccc	agagcttcc	ggccacctgc	gtgaacggcg	tgtgctggac	cgtgtaccac	180
ggcgccggca	caagaaccc	ggccggcccc	aaggcccc	tcaccagat	gtacaccaac	240
gtggaccagg	acctgtgtgg	ctggcaggcc	ccccccggcg	cccgagcct	gacccctctg	300
acctggcgca	cgagcgacct	gtacctgtgt	accgcgcacg	ccgacgtgat	ccccgtgcgc	360
cgccggcgcg	acagccggcg	cagcctgctg	agcccccgcc	ccgtgagcta	cttgaagggc	420
agcagcgccg	gccccctgct	gtgccccagc	ggccacgcgc	tgggcatctt	ccgcgcgcgc	480
gtgtgcaacc	gcggcgctgg	caaggccgtg	gacttctgtc	ccgtggagag	catggagacc	540
accatgcgca	gccccctgtt	caccgacaac	agcagcccc	ccgcctgtcc	ccagagcttc	600
cagggtggcc	acctgcacgc	ccccaacggc	agcggcaaga	gcaccaaggt	gcgcccgccc	660
tacgcgcgcc	aggtgtacaa	ggtgtctgtg	ctgaaccccc	gcgtggccgc	cacctgtggc	720
ttcgcgctct	acatgagcaa	ggccaccggc	atcgacccca	acatccgcac	cgcggtgcgc	780
accatcacca	ccggcgcccc	cgtagacctc	agcacctacg	gcaagtctct	ggccagagcg	840
ggctgcagcg	ggggcgccct	agcatcatc	atctgcgacg	agtgccacag	caccagacag	900
accacatccc	tgggcatcgg	caccgtgctg	gaccaggccg	agaccggcgg	cgcccgcctg	960
gtgggtgctg	ccaccggcac	ccccccggcg	agcgtgacgc	tgccccaccc	caactatcga	1020
gaggtggcccc	tgagcaaac	ggcgagatc	cccttctacg	gcaaggccat	ccccatcgag	1080
gcacatcccg	ggcgccgcca	ctgatctctc	tgccacagca	agaagaagtg	cgacagagctg	1140
gccgcgaagc	tgagcgccct	gggcatacac	gcccgtggct	actaccggcg	ctctggagctg	1200
agcgtgatcc	ccaccatcgg	cgacgtgtgt	gtgtgtggca	ccgacgcccc	gatgaccggc	1260
tacacggcgcg	cttctgacag	cgtgatcgac	tgcacaacct	cgctgacca	cgctgtggag	1320
ttacagcttg	ccccaccctt	caccatcgag	accaccacgc	tgccccagga	cgccgtggag	1380
cgacggcagc	cgccggcgcg	caccggcgcg	ggccgcgcgc	gcattctacg	cttctgtgac	1440
ccggcgagcg	gccccagcg	catgtctgac	agcagcgtgc	tgctcgagtg	ctacgacgac	1500

ggctgcgcct	ggtacgagct	gaccccgcc	gagaccagcg	tgcgcctcg	cgctacctg	1560
aacaccccc	gcctgcgcct	gtgcacaggac	caactggagt	tctgggagag	cggttccacc	1620
ggcctgaacc	acatcgagcg	ccacttccgt	agccagaccca	agcaggcccg	cgacaacttc	1680
ccctacctgg	tggcctacca	ggccaccgtg	tgcgcccgcc	cccagccccc	ccccccagc	1740
tgggaccaga	tgtggaagt	cctgatccgc	ctgaagccca	ccctgcacgc	cccccccccc	1800
ctgctgtacc	gcctggggcg	cgtgcgaacac	gaggtgaccc	tgacccaccc	catcaccaag	1860
tacatcatgg	ctctcatgag	cgcgcacctg	gaggtggtga	cagcacctgt	gggtgctggt	1920
ggcggcgctg	tggcgccct	ggcgcctac	tgcctgacca	ccggcagcgt	gggtatcggt	1980
tacatcatgg	ctctgagcgg	cgcgcgcgc	atcgtgcccg	accgcgagtt	ctgtaccag	2040
gagttcgacg	agatggaggga	gtgcgcacgc	caactgcctc	acatcgagca	gggcatgacg	2100
ctggccgagc	agttcaagca	gaaggccctg	ggcctgctgc	agaccgccac	caagcaggcc	2160
gagggcccg	ccccctggt	ggagagcaag	tggcgccgcc	tggagacctt	ctgggccaag	2220
cacatctgga	acttcatcag	cggcatccag	tacctggccg	gcctgagcac	cgctcccgcc	2280
aaccccccca	tgcgcagcct	gatggccttc	accgcagcac	tcaccagccc	ccctgacccc	2340
cagagcaccc	tgctgttcaa	catcctgggc	ggctgggtgg	cgcccgagct	ggcccccccc	2400
agcgccgcca	gcgcctctgt	ggcgcccgcc	atcgcccgcc	ccgcctgggg	cagcatcgcc	2460
ctgggcaagg	tgctgggtgga	catcctggcc	ggctacggcg	cggcgctggc	cgccgcctgt	2520
tggcctcata	tggatgatgag	cggcgagatg	cccagcacgc	aggacctggt	gaacctgctg	2580
cccgccatcc	tggcccccg	cgccttggtg	gtgggctgtg	tgtgcgccgc	catcctggag	2640
cgccacgttg	ggcccgcgga	ggcgcccggt	cagtggtgat	accgctgat	cgccctcgcc	2700
agcccgccga	accacgtgag	ccccaccac	tacgtgccgc	agagcgacgc	cgcgcgcgcg	2760
gtgacccaga	tcctgagcag	cctgaccatc	accagctgct	tgaagcgct	gcaccagtgg	2820
atcaacagag	actcgacac	ccccctgcag	ggcagctggc	tgcgcgacgt	gtgggacttg	2880
atctgcacgc	tgtgcacgga	cttcaagacc	tggctgcaga	gcaagctgct	gccccagctg	2940
cccgcgctgc	ctctcttcag	ctgcacgcgc	ggctacaagg	ggctgtggcg	cggcgacggc	3000
atcatcgaga	ccactgcctc	ctggcgccgc	cagatcaccc	gccacgtgaa	gaacgcgacg	3060
atgcgcacgc	tgggccccaa	gaactgcagc	aacacctggc	accgacacct	ccccctaac	3120
gcctacacca	cgggccccct	cacccccagc	cccgccccca	actacagccg	cgccctgtgt	3180
cgctgtggcg	ccgagggagta	cgtggagggt	accgcgtg	cgacttcca	tacgtgacc	3240
ggcatgacca	ccgacaacgt	gaagtgcctc	tggcaggtgc	ccgccccga	gtctcttacc	3300
gaggtggacg	gcgtgcgcct	gcacgcctac	gccccgcct	ccgccccct	gcctgcgcgag	3360
gaggtgacct	tccaggtggg	cctgaacacg	tacctgggtg	gcagccagct	gcctgcgag	3420
cccgagcccg	acgtggccgt	gctgaccagc	atgtgacgc	accccagcca	catcacccgc	3480
gagacggcca	agcgccgcct	ggcccgccgc	agccccccca	gcctggccag	cagcagccgc	3540
agccagctga	ggccccccag	cctgaaggcc	accctgacca	cccaccagct	gaaccccgac	3600
gcgcacctga	tggaggccaa	cctgctgtgt	cgccaggaga	tggcgcccaa	catcacccgc	3660
gtggagagcg	agaacaagg	ggtgtgtctg	gaacgcttcg	acccctgcgc	cgccgaggag	3720
gcgcagcgcg	aggtgagcgt	ggccgcgag	atcctcgcca	agagcaaaag	gtcccccgct	3780
gcacatccca	tctgggctag	acctgattac	aacctccccc	tgtggagag	ctggaagac	3840
ctgattacg	tgccctccag	ggcgatggc	tgtctctctg	ctccactaa	agccctctct	3900
atccaccttc	ctaggcgcaa	aaggcacctg	gtgctgacag	aaagcgagct	gagctctgct	3960
ctggcggaac	ggccacacaa	gaaccttggc	agcagcgaga	gcctccgct	ggcagcggga	4020
acagccacccg	ctctgctgta	caggccacgc	gacgacggcg	ataagggcag	cgaagtggag	4080
agctatagca	gcattgctcc	cctggaaagg	gaacctggcg	atcccgatct	gaacgatggc	4140
agctggagca	ccgtgagcga	agaggccacg	gagagcgtgg	tgtgttgcag	catgagctac	4200
acctggacac	gcctctgat	acacacctgc	gctgcggagc	agagcaagct	ggccatcaac	4260
gccttgagca	acagctctgt	gaggcacacc	aaatggtgtg	accgccacac	cagcaggtct	4320
gcggcgatga	ggcagaagaa	ggtagacctc	gaacctctgc	aggtgcttga	gcacacctac	4380
cgcgatgtgc	tgaaggagat	gaaggccaa	gccagcagct	tgaaggccaa	gctgctgagc	4440
gtggaggagg	cttgcaagct	gacccccccc	cacagcgcca	agagcaagtt	gaacgaggtg	4500
gccaaaggag	tggcgaacct	gagcagcaag	gccgtgaacc	acatccacag	cgtgtggagag	4560
gaactctcgg	aggaacacct	gaacccccatc	gaacaccaca	tcatggccaa	tcaggagggtg	4620
ttctgcgtgc	agcccagaaa	ggcgcccgcc	aagccccccc	gcctgatcgt	gtcccccgac	4680
ctggcgctgc	gcgtgtgcga	gaagatggcc	ctgtacgacg	tgttgagcac	cttgcgccag	4740
gtggtgatgg	gcagcagcta	cggctccag	tacagccccc	gccagcgctg	ggagtctcctg	4800

```

gtgaacacct ggaagagcaa gaagaacccc atggggttca gctacgacac cggctgcttc 4860
gacagacacc tgaccgagaa cgacatccgc gtggaggaga gcatctacca gtgctgcgac 4920
ctgggccccg aggcccgcca ggcatcaag agcctgaccg agcgctgtga catcgccggc 4980
cccttgacca acagcaaggg ccagaactgc ggctaccgcc gctgcgcgcg cagcggcgtg 5040
ctgaccacca gctcgggcaa acccctgacc tgctacctga aggccagcgc cgctgcccgc 5100
ggccgcaagc tgcaggactg caccatgctg gtgaacgccg ccggcctggt ggtgatctgc 5160
ggagggcgcg gcacccgaga ggacgcgcgc agcctgcgcg tgttaccaga ggccatgacc 5220
cgctacagcg cccccccggg cgaccccccc cagcccgagt acgacctgga gctgatcacc 5280
agctgcagca gcaactgtag cgtggcccac gacgccagcg gcaagcgtgt gtactacctg 5340
accgcgacc ccaccacccc cctggcccgc gcgcgctggg agaccgcccg ccacaccccc 5400
gtgaacagct ggtggggcaa catcatcatg tacgcccaca cctgtggggc cgcgatgac 5460
ctgatgacc acctcttcag catcctgctg gcccaggagc agctggagaa ggccttggac 5520
tgccagatct acggcgctcg ctacagcctc gagcccttgg acctgcccca gatcacggag 5580
cgccctgcacg gcttgagcgc ctccagcctg cacagctaca gcccccggga gatcaaccgc 5640
gtggccagct gcttcgcgaa gctgggcgtg cccccctgc gcgtgtggcg ccaccgcgcc 5700
cgacagcgtc ggcgccgctt gctgagccag ggcggccgcg ccgccacctg cggcaagtac 5760
ctgttcaact gggcgtgtga gaccaagctg aagctgaccg ccacccccgc cgccagccag 5820
ctggacctga cggcgtgtgt cgtggccggc tacagcggcg gcgacatcta ccacagcctg 5880
agccgcgccc gcccccgtcg gttcatgctg tgccctgctg tgctgagcgt gggcgctggg 5940
atctactctg tgcccaaccg ctaaa

```

<210> 12

<211> 10

<212> RNA

<213> Artificial Sequence

<220>

<223> Ribosome binding site

<400> 12

gccaccaugg

10

<210> 13

<211> 49

<212> RNA

<213> Artificial Sequence

<220>

<223> Synthetic polyadenylation signal

<400> 13

aauaaaagau cuuuuuuuuc auuagaucug uguguugguu uuuuugugug

49

<210> 14

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Additional nucleotides present in pVIJns-NS

<400> 14

tctagagcgt ttaaaccttt aattaagg

28

<210> 15

<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Additional nucleotides present in pV1Jns-NSOPTmut

<400> 15
tttaaagtgtt taaac 15

<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide primer

<400> 16
tcgaatcgat acgcaaacct acgc 24

<210> 17
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide primer

<400> 17
tcgacgtgtc gacttcgaag cgcacaccaa aaacgtc 37